

Arbetsrapport nr 62

En introduktion till logistisk regressionsanalys

Johannes Bjerling
Jonas Ohlsson



GÖTEBORGS UNIVERSITET

Institutionen för journalistik, medier och kommunikation

Arbetsrapport nr. 62

En introduktion till logistisk regressionsanalys

Johannes Bjerling
Jonas Ohlsson

ISSN 1101-4679



GÖTEBORGS UNIVERSITET

Institutionen för journalistik, medier och kommunikation

Box 710, 405 30 GÖTEBORG

Telefon: 031-786 49 76 • Fax: 031-786 45 54

E-post: majken.johansson@jmg.gu.se

2010

Innehåll

Inledning	1
1.1 Vad är logistisk regression?	2
1.2 Varför odds? Och varför logaritmer av odds?	3
1.3 Likheter och skillnader, fördelar och nackdelar	4
1.4 Vårt exempel: Vad påverkar förtroendet för politiker?	5
2.1 Vi tar oss an vårt material!	7
3.1 Olika signifikanstester	10
3.2 Standardfelfamiljen	10
3.2.1 Z-test	10
3.2.2 Wald-test	11
3.3 Likelihood ratio test	12
3.4 Wald eller Likelihood ratio – vad är bäst?	19
3.5 Hur ”bra” är modellen? Hosmer och Lemeshow ger svar	20
3.6 Relevansen av R^2	23
4.1 Resultaten presenteras	24
4.2 Övrig redovisningsformalia	27
5.1 Sammanfattning och slutsatser	28
6.1 Referenser	30

Inledning

Inom naturvetenskapen arbetar man ofta med data som befinner sig på intervall- eller kvotskalnivå. En biolog kan till exempel vilja veta om en välgödd gås lägger större ägg än en mager gås, medan en fysiker istället intresserar sig för förhållandet mellan avstånd och kraft. Rena, kvantitativa data. Inga brott eller gränser, lite som när man efter ett djupt andetag blåser upp en ballong: volymen ökar kontinuerligt. (Och eftersom vår sansade naturvetare vet när det är dags att sluta blåsa förutsätter vi att ballongen inte kommer att spricka!)

Inom samhällsvetenskapen är det mindre vanligt att man intresserar sig för äggstorlek och naturlagar. Istället bygger samhällsvetenskapliga data ofta på värden som är kvalitativa. Att klass och kön (s.k. nominaldata) är kvalitativa är kanske uppenbart, men att desamma gäller för olika sorters rangordningsdata (ordinaldata) kan vara svårare att se.

Trots skillnaderna i data är emellertid ofta målet detsamma: med hjälp av teorier och data söker såväl naturvetaren som samhällsvetaren förklaringar till vår omvärld. Varför stenen faller ska på samma sätt som ett sjunkande valdeltagande förklaras.

Men även om målet är detsamma finns det inget som säger att vägen bör vara densamma. Även om statistiska metoder kan vara användbara för såväl natur- som samhällsvetare kan valet av statistisk metod se olika ut. Eller, mera korrekt: ofta *bör* valet av metod se olika ut. Vad det handlar om är att anpassa den statistiska analysmetoden efter materialet.

Och det är av just den här anledningen som vi menar att logistisk regression är en metod som förtjänar att lyftas fram. Tre aptitretare kan presenteras redan här:

För det första: I en (binominal) logistisk regression går det utmärkt att arbeta med kvalitativa data, den beroende variabeln är *binär*.

För det andra: Eftersom den beroende variabeln är binär finns inga normalfördelningskrav, metoden kan hantera en *snedfördelad datamängd*.

För det tredje: Relationen mellan de oberoende variablerna och den beroende variabeln behöver inte vara linjär, metoden fungerar på *kurvlinjära samband*.

Tanken med det här kompendiet är att introducera logistisk regression på ett så lättillgängligt sätt som möjligt. Detta innebär att texten kan komma att göra den som är intresserad av en mer teoretisk förståelse av metoden besviken, snarare än att redogöra för matematiska formler är vår avsikt att fokusera på frågor och problem ur ett praktiskt perspektiv.

In essence, the logistic model predicts the logit of Y from X. [...] The logit is the natural logarithm (\ln) of odds of Y, and odds are ratios of probabilities (π) of Y happening (i.e., a student is recommended for remedial reading instruction) to probabilities ($1-\pi$) of Y not happening (i.e., a student is not recommended for remedial reading instruction).

Peng, Lee & Ingersoll 2002:4

1.1 Vad är logistisk regression?

I en utmärkt introduktion till metoden skriver Per Arne Tufte (2000:7f) att logistisk regression är "[e]n metode for å behandle kvalitative, avhengige variabler ... Fra å være relativt lite brukt på begynnelsen av 90-tallet, er den i dag nesten den dominerende formen for regresjonsanalyse innen sosiologisk og statsvitenskaplig forskning. En av de viktigste grunnene til dette er naturligvis at langt de fleste avhengige variabler disse fagdisiplinene arbeider med er dikotome eller på nominal-/ordinalnivå."

Låt vara att Tuftes text snart har tio år på nacken, logistisk regression *är* en metod på framfart. Och, som Tufte också skriver, en av förklaringarna är att logistisk regression fungerar utmärkt också för kvalitativa data.

Men varför har då dess genombrott dröjt? Metoden har ju funnits sedan 1960-talet slut (Cabrera 1994). Svaret torde kunna skrivas med ett enda tecken: e .¹ Många samhällsvetare ryggar förmodligen tillbaka av att den beroende variabeln är den *naturliga logaritmen* av ett odds – det gjorde åtminstone vi. Men saken är den att man faktiskt inte behöver vara särskilt matematiskt bevandrad för att praktiskt arbeta med logistisk regression. Accepterar man att den naturliga logaritmen av ett tal (a) är det tal (x) som e ska upphöjas till för att ge (a), ja då är man hemma.

Vi tar det en gång till, för att sedan lämna talet e helt och hållet. Den naturliga logaritmen av ett tal (a) är det tal (x) som e ska upphöjas till för att ge (a); den naturliga logaritmen av 10 är alltså det tal som e ($\approx 2,72$) ska upphöjas till för att ge 10. Den naturliga logaritmen av 10 är med andra ord 2,30 ($\ln 10 \approx 2,30$).

En annan sak som gör att logistik regression skiljer sig från vanlig OLS-regression är att den beroende variabeln utgörs av ett odds, alltså sannolikheten för ett utfall dividerat med dess motsats. Om sannolikheten (p) för borgerliga sympatier är 60 procent kommer oddset (o) för borgerliga sympatier att vara 1,5 ($60/40 = 1,5$). Tar vi den naturliga logaritmen av oddset får vi den så kallade logiten. Det är logiten som är vår beroende variabel. Oddset 1,5 ger logiten 0,41 ($\ln 1,5 \approx 0,41$).

Innan vi går igenom finessen med logaritmerade odds sammanfattar vi:

- Logistisk regression fungerar för alla sorters data.
- Den beroende variabeln uttrycks som ett odds.
- När oddset logaritmeras ($\ln o$) får vi den så kallade "logiten".

¹ Eulers tal, det mystiska talet e . 2.718281828459045235360... och så vidare, i oändlighet. (För den som är intresserad finns de första 100 000 decimalerna på följande adress <http://www.mu.org/~doug/exp/100000.html>).

1.2 Varför odds? Och varför *logaritmer* av odds?!

Ovan har det framgått att en påtaglig skillnad mellan logistisk regression och vanlig OLS är att man i logistisk regression genomgående jobbar med odds. Eftersom oddset uttrycker sannolikheten genom formeln $o = p/1-p$ kommer en *stor* sannolikhet att motsvaras av ett *högt* odds. För den som är van vid hur odds uttrycks i samband med spel kan detta verka lite märkligt, på travbanan innebär ju en högoddsare det omvända!

Vårt svar till detta är kort men kärnfullt: Allt här i livet är inte logiskt. Ett högt odds där är ett lågt odds här. Låt oss istället angripa en väsentligare fråga: Vad är *vitsen* med att jobba med odds? Och kan man inte komma undan det här med logaritmer?

Begrunda för ett ögonblick tabellen som presenteras nedan!

Tabell 1.1 Förhållandet mellan sannolikheter, odds och logaritmerade odds

Sannolikhet (p)	Odds (o)	Logaritmerade odds (logit)
25 %	0,333	-1,099
50 %	1,000	0
75 %	3,000	1,099
90 %	9,000	2,197

Av tabellen framgår att en sannolikhet på 25 procent ger oddset 0,33 (25/75); 50 procents sannolikhet motsvaras av oddset 1 (50/50), medan 75 procents sannolikhet ger oddset 3 (75/25) och 90 procents sannolikhet ger oddset 9 (90/10). För varje rad som vi rör oss uppåt/nedåt förändras alltså *oddset* med faktorn 3.

Av kolumnen längst till vänster kan vi samtidigt konstatera att den procentuella ökningen planar ut – från 25 procentenheter till 15 procentenheter. Skulle vi multiplicera oddset med 3 en gång till blir utplaningen ännu tydligare: Medan oddset ökar till 27 (3*9) ökar sannolikheten med bara 6 procentenheter ($o = 27$ motsvaras av $p = 96$ procent). Ytterligare en multiplicering med 3 ger oddset 81 (3*27), medan sannolikheten bara ökar med 3 procentenheter ($o = 81$ motsvaras av $p = 99$ procent).²

När vi jämför mittkolumnen med kolumnen till höger upptäcker vi ytterligare en intressant liten detalj: Mittkolumnens kurvlinjära samband är linjärt. För varje gång som oddset multipliceras med 3 ökar logiten med 1,099. Från -1,099 hamnar vi i tur och ordning på 0, 1,099 och 2,197. Går vi ytterligare en rad nedåt kommer logiten att bli 3,296 (2,197 + 1,099). Genom att oddsen logaritmeras transformeras det kurvlinjära till att bli likformigt. Om vi presenterar värdena som vanliga tal eller som logaritmerade tal kommer alltså att ha en direkt inverkan på vilken form funktionen kommer att anta.³

Finessen med logaritmerade odds kan sammanfattas i två korta meningar: 1) Att sannolikheten uttrycks som odds innebär att den *övre* gränsen elimineras. 2) Att oddsen logaritmeras innebär att den *nedre* gränsen elimineras. Variabeln kommer därmed att uppfylla

² Motsvarande gäller givetvis också om vi går åt motsatt håll, alltså mot ett lägre odds. Uttryckt i procentenheter kommer förändringen gradvis att bli allt mindre.

³ Jämför med relationen mellan vanliga tal och dess "tiologaritmer": $1000 = 10^3$; $100 = 10^2$ och $10 = 10^1$. Betraktas talen som tiologaritmer beskriver de en linjär funktion; betraktas de som vanliga tal beskriver de en kurvlinjär funktion.

en viktig förutsättning för OLS-regression: Genom att värdena anges som logaritmerade odds blir variabeln oändlig.

En intuitiv förståelse av S-kurvan: Föräldraledighet som en funktion av antalet barn

I vanlig OLS-regression antas sambandet mellan oberoende och beroende variabler vara linjärt. Innebörden av detta kan formuleras som "oavsett var på x-axeln vi befinner oss antas en viss variabel ge upphov till en viss effekt". Men är det verkligen rimligt att anta att ett stimuli har en konstant effekt? Invändningen blir begriplig om vi använder oss av ett exempel!

Frågeställningen som intresserar oss är hur antalet barn påverkar sannolikheten att en person är föräldraledig. Vår oberoende variabel (x) är alltså antalet barn, vår beroende variabel (y) är sannolikheten för föräldraledighet.

Att det i det här fallet knappast handlar om en konstant effekt behöver man inte vara ett geni för att begripa, sannolikheten för föräldraledighet ökar snabbt mellan 0 och 1 för att sedan plana ut. För föräldraledighet gör alltså "0,5 barn" varken till eller från, föräldraledighet tycks närmast förutsätta föräldraskap (vilket alla, föräldrar eller ej, nog anser som rimligt). Med logistisk regression kommer vi förbi problemet, genom att oddsen logaritmeras blir det kurvlinjära linjärt!

Innan vi går vidare sammanfattar vi:

- Odds innebär att "taket" försvinner, logaritmerade odds innebär att "golvet försvinner".
- Genom att sannolikheten uttrycks som logaritmerade odds blir S-kurvan linjär. Det är de logaritmerade oddsen som gör att vi kan tala om "en konstant effekt" utan att göra våld på verkligheten!

1.3 Likheter och skillnader, fördelar och nackdelar

Vi börjar med de mest uppenbara likheterna: De logaritmerade koefficienterna har sin motsvarighet i OLS-regressionens ostandardiserade b-koefficienter och precis som i OLS kan syftet för regressionen vara att jämföra enskilda variabler eller hela modeller. (Även om meningsfullheten kan diskuteras så *går* det att få fram en form av pseudo R^2 , mer om detta senare!)

Samtidigt som likheterna överväger finns det emellertid också skillnader. Den utifrån ett praktiskt perspektiv mest påtagliga skillnaden – att den beroende variabeln utgörs av logaritmerade odds – har behandlats ovan, men också *förutsättningarna* ser delvis olika ut. Logistisk regression bygger t.ex. *inte* på att sambandet är linjärt (se ovan) och kravet på normalfördelning är upphävt. Jämförs villkoren för logistisk regression med de krav som ställs i samband med OLS-regression kan man – inte utan viss lättnad – konstatera att logistisk regression överlag är en betydligt mera tillåtande metod (jmf Garson, "Statnotes").

Att metoden, trots denna uppenbara fördel, inte förekommer oftare torde främst förklaras av den något omständliga procedur som krävs för att resultaten skall bli begripliga. Vad koefficienterna anger är, till exempel, förändringen på den *logaritmerade* oddskvoten (logiten). För att bli användbara måste koefficienterna därför "avlogaritmeras" – och riktigt tydliga blir resultaten endast då de omvandlas till predicerade sannolikheter.

För att metoden ska vara ett effektivt och användbart verktyg krävs det alltså ett gediget teoretiskt förarbete. Vilket samband är det som vi är intresserade av att pröva? Vilken riktning tänker vi oss att sambandet har? Och när det gäller den beroende variabeln: Är det sannolikheten för att ett visst fenomen ska *inträffa* som studeras? Eller är vi tvärtom intresserade av sannolikheten för att något *inte* inträffar?

Innan vi kastar oss över vårt exempel sammanfattar vi hur oddskvoten ska förstås:

- Oddskvoten visar hur många gånger oddset förändras av ett skalsteg på den oberoende variabeln.
- Oddskvoter som är större än 1 innebär att oddset ökar, medan oddskvoter som är mindre än 1 innebär att oddset minskar.

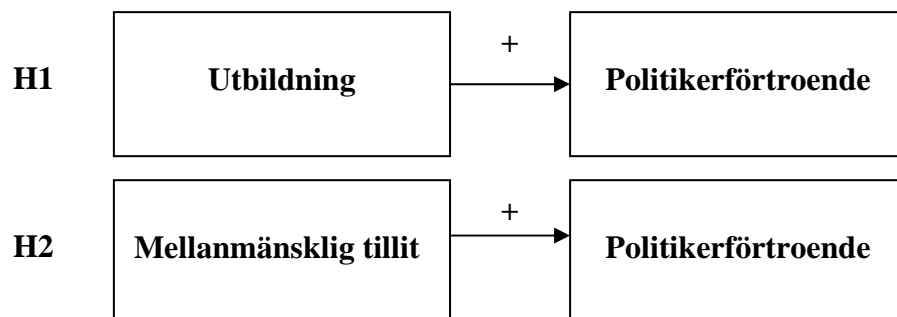
Observera: Vill vi veta hur oddset förändras av exempelvis fyra skalsteg på den oberoende variabeln kan vi *inte* multiplicera oddskvoten med fyra. För att få med den stegvisa förändringen måste kvoten istället *kvadreras* med fyra! (Jämför med ränteutvecklingen på sparade pengar: Om det insatta beloppet är 1 000 kr och räntesatsen är 3 % kommer saldot efter tio år att visa $1,03^{10} * 1\,000\text{ kr} = 1\,343\text{ kr}$. Om räntan *inte* hade förökats hade vi fått nöja oss med $10 * 1,03 * 1\,000\text{ kr} = 10\,300\text{ kr}$.)

1.4 Vårt exempel: Vad påverkar förtroendet för politiker?

Eftersom vår ambition har varit att introducera metoden ur praktiskt perspektiv har vi valt att bygga presentationen kring ett konkret exempel. Data kommer från 2007 års SOM-undersökning, nedan redogörs för de variabler vi har valt att arbeta med samt de antaganden som ligger bakom den gjorda analysen.

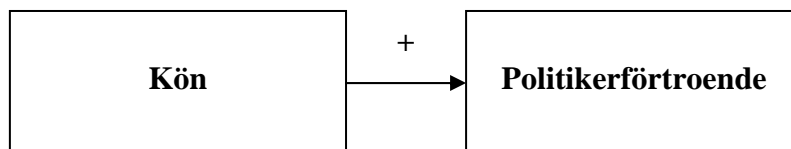
- Oberoende variabler: utbildning, kön och mellanmänsklig tillit
- Beroende variabel: sannolikheten för stort politikerförtroende

Av de oberoende variablerna är utbildning och mellanmänsklig tillit på ordinalskaleivå. I båda fallen förmodar vi att det finns positiva samband med den beroende variabeln; sannolikheten för ett stort politikerförtroende antas *öka* med både utbildningsnivå och mellanmänsklig tillit.



För den tredje av våra oberoende variabler, kön, har tidigare SOM-undersökningar visat att förtroendet för politiska institutioner är högre bland män än bland kvinnor. Detta ger oss vår tredje hypotes: Politikerförtroendet antas vara större bland män än bland kvinnor.

H3



Eftersom vi vill att riktningen i samtliga fall skall vara densamma försäkras vi oss om att kvinna är kodat som 0 och man som 1.

Variabeln som ligger till grund för vår beroende variabel, sannolikheten för ett stort politikerförtroende, är ursprungligen fyrdelad. Variabelkonstruktionen gör emellertid dikotomiseringen tämligen oproblematisk: 1 (mycket stort förtroende) och 2 (ganska stort förtroende) slås samman och bildar en kategori för stort förtroende; 3 (ganska litet förtroende) och 4 (mycket litet förtroende) slås samman och bildar en kategori för litet förtroende.

Den kategori som intresserar oss (stort förtroende) ger vi värdet 1 (litet förtroende = 0). Med denna kodning kommer körningarna att visa hur sannolikheten för stort politikerförtroende påverkas av våra oberoende variabler.

För att lära känna vårt material före det att vi utsätter det för en logistisk regressionsanalys gör vi först en enkel genomgång av hur de enskilda förklaringsvariablerna förhåller sig till den oberoende variabeln.

Tabell 1.2 Andelen svenskar med stort förtroende för svenska politiker, efter utbildning, kön och mellanmänsklig tillit, 2007 (procent)

Samtliga											
%	34										
n	1 620										
Utbildning	Låg	Medellåg		Medelhög		Hög					
%	24	30		39		47					
n	357	532		329		373					
Kön	Kvinna	Man									
%	30	39									
n	862	758									
Mellanmänsklig tillit (0-10)	0	1	2	3	4	5	6	7	8	9	10
%	10	0	26	21	25	28	36	33	41	48	47
n	29	19	51	86	91	251	133	325	365	137	95

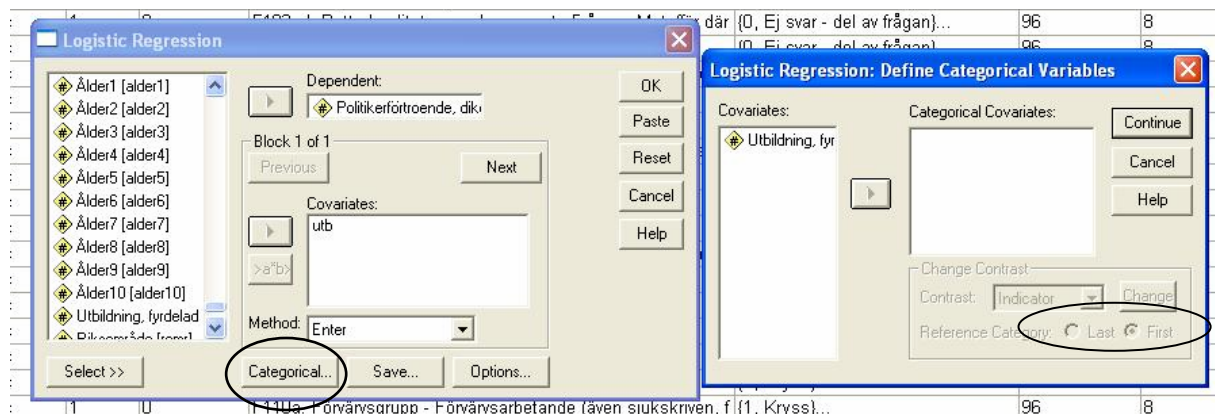
Kommentar: Data från den nationella SOM-undersökningen 2007. Den fyrgradiga svarsskalan till frågan om förtroende för svenska politiker har dikotomiserats ("mycket litet" samt "ganska litet" förtroende = 0; "ganska" samt "mycket stort" förtroende=1). I figuren visas andelen med stort förtroende (1).

Vi ser i tabellen att andelen svenskar som angav sig ha antingen ganska eller mycket stort förtroende för svenska politiker år 2007 uppgick till 34 procent. Resterande 64 procent hade följaktligen antingen ganska eller mycket litet förtroende för våra folkvalda representanter. Härutöver ser vi att andelen med stort förtroende stiger med såväl utbildning, som kön (från kvinna till man) och mellanmänsklig tillit. Det tycks således finnas goda skäl att gå vidare med arbetet med våra hypoteser. Frågan är vad som händer när analysen av de enskilda variablerna görs med kontroll för varandra?

2.1 Vi tar oss an vårt material!

Eftersom utbildningsvariabeln är fyrdelad och på ordinalskalenivå väljer vi att först av allt dummykoda variabeln.⁴ Jobbar man i SPSS är det bara att tacka och ta emot: Omvandlingen görs av programmet! Allt man behöver göra är att i menyfönstret ange

- 1) att det handlar om en *kategorisk variabel* och
- 2) om man vill att referenskategori ska utgå från det sista eller det första värdet. Last och First blir valbara genom att variabeln med piltangentens hjälp flyttas till det högra fönstret (Categorical Covariates). Vi väljer att använda det första värdet (1, låg utbildning) som referenskategori.



Och sedan är det bara att trycka på Continue! Om allt har gått rätt ska utbildningsvariabeln ha bytt namn från utb till utb(Cat). Efter att ha fört in våra andra oberoende variabler är det bara att klicka OK – körningen kan göras...

I den output som sedan presenteras börjar vi med tabellen längst ned (Variables in the Equation). Likheten med OLS-regression är slående. Precis som i OLS visar den första kolumnen våra oberoende variabler och en konstant.

Av en för oss okänd anledning kallas de oberoende variablerna för "Covariates". Eftersom vår hållning är pragmatisk väljer vi att inte ödsla tid och energi för att ta reda på varför – så länge saker och ting funkar nöjer vi oss med det!

Utbildningsvariabeln, som vi valde att göra om till en dummy, består nu av UTB(1), UTB(2) och UTB(3). Eftersom referenskategori är låg utbildning kommer UTB(1), UTB(2) och UTB(3) i tur och ordning att stå för medellåg, medelhög och hög utbildning.

⁴ I likhet med OLS-regressionen förutsätter den logistiska regressionsanalysen att det är lika långt mellan skalstegen på den oberoende variabeln (jfr intervall- och kvotskalor). Nu går det att använda även ordinal- och nominalskalor, men då måste variablerna först dummykodas, vilket innebär att variabelvärdena dikotomiseras (t.ex. man = 0, kvinna = 1). Om variabeln har fler värden än två måste fler dummyvariabler skapas. När så är fallet måste man först bestämma sig för en referenskategori, som sedan aldrig kodas som 1 i egen dummy, utan endast ingår som 0-värde hos övriga dummyvariabler – detta för att undvika s.k. multikolinjaritet, dvs. (alltför) hög samvariation mellan två variabler (för mer om dummyvariabler och dummykodning, se Garson 2009)

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a			40,006	3	,000	
utb						
utb(1)	,311	,163	3,642	1	,056	1,364
utb(2)	,760	,175	18,771	1	,000	2,139
utb(3)	,952	,171	31,060	1	,000	2,592
Kön	,544	,112	23,705	1	,000	1,722
F106a	,170	,027	39,196	1	,000	1,185
Constant	-2,550	,228	125,236	1	,000	,078

a. Variable(s) entered on step 1: F106a.

I kolumn 2 (B) visas de ostandardiserade koefficienterna. I den form som de *här* presenteras vore det lögn att säga att värdena är intuitivt begripliga: Vad som anges är den *naturliga logaritmen av oddsets förändring* – och eftersom ”de färreste av oss tenker naturlig i termer av log-odds” (Tuft 2000:26) nöjer vi oss här och nu med ett enkelt konstaterande: positiva värden innebär positiva samband. Oddset (och därmed också sannolikheten) *ökar* alltså med stigande värden på den oberoende variabeln.

Ser vi till utfallet i tabellen så tycks våra tre hypoteser få stöd i analysen: Samtliga oberoende variabler har positiva koefficienter, stiger värdena på våra oberoende variabler kommer alltså sannolikheten för stort politikerförtroende att öka. (Negativa B-värden hade inneburit det omvända; med högre värden på våra oberoende variabler skulle sannolikheten ha minskat.)

I kolumn 3 (S.E) hittar vi – precis som vid OLS-regression – standardfelen till koefficienterna, och i de tre följande kolumnerna (Wald, df, Sig) återfinns resultaten av signifikansprövningarna. Vi kommer tillbaka till tester av signifikans senare.

Vad vi redan nu kan se är att samtliga variabler utom en är signifikanta: Variabeln UTB(1) – d.v.s. medellåg utbildning – är inte statistiskt signifikant från referensgruppen.

I den sjunde och sista kolumnen (Exp[B]) presenteras de anti-logaritmerade *oddskvoterna*. Eftersom värdena anger kvoten av det nya oddset dividerat med det gamla oddset (se ovan) kan de redovisade värdena uppfattas som procent. För att tydliggöra vad vi menar tittar vi närmare på värdena i den näst översta raden.

I kolumn 2 har den ostandardiserade koefficienten för medellåg utbildning värdet 0,311. Den tillhörande anti-logaritmen i kolumn 7 är 1,364. Innebörden är att oddset ökar 1,364 gånger när vi går från låg utbildning till medellåg utbildning (när övriga oberoende variabler konstanthålls vid respektive medelvärde). Att oddset ökar 1,364 gånger innebär att oddset har blivit 36,4 procent större.

Ekvationen för att beräkna "logiten" – och för att sen få fram den faktiska sannolikheten!

Ovan har det framgått att koefficienterna i kolumn två anger det logariterade värdet av oddsets förändring. Själva logiten ($\ln p/(1-p)$) redovisas alltså inte, vad körningen visar är hur de oberoende variabelnas påverkar vårt logariterade odds. Logiten (z) beräknas enligt formeln

$$z = \ln(\text{odds}_{\text{utfall}}) = \ln(\text{sannolikhet}_{\text{utfall}}/\text{sannolikhet}_{\text{motsatt utfall}}) = \ln(\text{sannolikhet}_{\text{utfall}}/[1 - \text{sannolikhet}_{\text{utfall}}]) =$$

$$b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

Bökigare än önskat? Ja, onekligen! Men exemplifierar vi med värden från tabellen ovan blir det klarare. Säg att vi vill veta logiten för en man med medellång utbildning och värdet 4 på variabeln som avser mellanmännisklig tillit. Värdena som då ska adderas blir

-2,550 (=konstanten) + 0,311*1 (=medellång utbildning) + 0,544*1 (=man) + 0,170*4 (mellanmännisklig tillit =4)

När de olika värdena slås samman får vi värdet -1,01698 som alltså är logiten (=ln o). För att få fram oddset anti-logariteras värdet, varpå vi får talet 0,362. Detta är alltså oddset för att personen ifråga skall ha stort förtroende för politiker. Oddset (o) kan lätt omvandlas till sannolikhet (p) då $p = o/(1+o)$.

Sannolikheten för att en man med medellång utbildning och värdet 4 på variabeln "mellanmännisklig tillit" ska ha stort förtroende för politiker är alltså 26 procent ($0,362/(1+0,362)$). Visst är metoden finurlig?!

"Att vara eller inte vara" – den beroende variabeln i logistisk regression

Skillnaden i hur den beroende variabeln ska tolkas blir tydlig om vi för in variablerna ovan i en OLS-regression. Outputen skulle då se ut som nedan!

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	-,025	,041		,612
	utb1 dummy	,057	,032	,057	,074
	utb2 dummy	,154	,036	,132	,000
	utb3 dummy	,203	,035	,181	,000
	Kön	,114	,023	,119	,000
	F106a	,033	,005	,158	,000

a. Dependent Variable: f46_ny Pol förtroende - dikotom

Längst till vänster hittar vi de ostandardiserade b-koefficienterna; i kolumnerna till höger anges i tur och ordning standardfelen, de standardiserade betakoefficienterna, t-värdena och signifikansnivåerna.

För den som är van vid OLS-regression är tolkningen av b-koefficienterna en baggis: Går vi ett skalsteg på respektive oberoende variabel så ökar den beroende variabeln med 0,057, 0,154 och 0,203 för de tre dummyvariablerna, och 0,114 respektive 0,033 för kön respektive mellanmännisklig tillit.

I logistisk regression är den beroende variabeln "relativ". Vad som anges är förändringen av en kvot, hur stor den "absoluta" förändringen är beror på det ursprungliga värdet!

Inte alldeles enkelt? OK, låt oss anta att den ursprungliga sannolikheten är 50 procent. Oddset är då 1 (50/50). Efter ett skalstegs ökning i den oberoende variabeln har sannolikheten stigit till 75 procent. Oddset är då 3 (75/25). Dividerar vi det "nya" oddset med det "gamla" får vi en "oddskvot" (som i det här fallet skulle vara 3). Att effekten är relativ är alltså en konsekvens av att kvoten uppstår då det nya oddset divideras med det gamla!

Innan vi går vidare olika signifikanstest sammanfattar vi:

- I SPSS kallas de oberoende variablerna för "Covariates", dummykodningen kan beställas eller göras manuellt.
- Precis som i OLS-regression visar koefficienterna effekten av ett skalsteg på våra oberoende variabler.
- De anti-logaritmerade oddskvoterna ($\text{Exp}[B]$) anger oddsets förändring och kan tolkas i termer av procent.
- Om logiten omvandlas till procent kringgås det snåriga i att sannolikheten uttrycks som odds. En enkel beräkning kommer alltså att göra resultaten betydligt lättare att tolka!

3.1 Olika signifikanstester

Arbetar man med urvalsundersökningar aktualiseras frågan om resultatens signifikans: I vilken mån riskerar ett samband som man eventuellt hittar att bero på slumpen? Och vice versa: I vilken mån riskerar ett faktiskt samband att framstå som slumpmässigt?

Det finns till den logistiska regressionen tre olika verktyg för att utföra signifikanstest av den s.k. nollhypotesen ($H_0: B = 0$), dvs. huruvida det finns något statistiskt samband mellan en beroende och en eller flera oberoende variabler. Som vi ska se finns här både skillnader och likheter jämfört med signifikanstester inom OLS-regressionen.⁵

3.2 Standardfelsfamiljen

För att pröva om ett resultat är signifikant eller inte måste man först beräkna ett mått på osäkerhet för de resultat vi får fram. Standardfelet är just ett sådant mått. Det anger i vilken grad värdet på en koefficient riskerar att avvika från det värde vi vill kunna uttala oss om, dvs. värdet för hela den studerade populationen. Med utgångspunkt i standardfelet finns det två huvudsakliga signifikanstester att välja bland: z-testet och Wald-testet.

3.2.1 Z-test

Vid logistisk regression går det att beräkna s.k. asymptotiska standardvärden. Dessa värden presenteras av SPSS i kolumnen ASE (för asymptotic standard errors). Huruvida en koefficient, β , är signifikant eller inte kan testas på samma sätt som ett z-test i OLS-regressioner⁶:

$$z = \beta / \text{ASE}$$

Låt oss säga att vi vill testa en femprocentig signifikansnivå: Om z är större än 1,96 kan vi förkasta nollhypotesen. Med 95 procents säkerhet kan vi alltså säga att det finns ett signifikant

⁵ SPSS redovisar även ett ytterligare, mindre vanligt förekommande, test, det s.k. score-testet. För den som är intresserad finns testet beskrivet av bl.a. Hosmer & Lemeshow (1989:17f) och Long (1997:87ff).

⁶ z-testet är ett signifikanstest som kan användas när populationens standardavvikelse och medelvärde är kända. z-testet kan anses vara ungefärligen normalfördelat när urvalet är stort och $\beta = 0$.

dubbelriktat⁷ samband mellan de båda variablerna. Som synes är principen precis den samma som vid vanliga signifikanstester av koefficienter i OLS.⁸

3.2.2 Wald-test

Så kommer vi då till det Wald-test, som den observante läsaren kommer ihåg från outputfönstret från SPSS i avsnitt 2.1. Wald-testet är ett dubbelriktat hypotestest som bygger på samma princip som ett vanligt z-test. Wald-testet beräknas helt enkelt genom att z-värdet kvadreras:

$$\text{Wald} = z^2 = (\beta / \text{ASE})^2$$

Resultat över det kritiska värdet 3,84 (dvs. kvadraten av 1,96; jfr ovan) innebär vid en femprocentig signifikansnivå att nollhypotesen kan förkastas. För en enprocentig signifikansnivå (0,01) gäller följaktligen 6,66 (dvs. 2,58²), och för 0,001-nivån 10,82 (3,29²). Wald-testet ger alltså precis samma resultat som z-testet.

Här nedan har vi återigen resultatet av vår logistiska regressionsanalys. Det har nu blivit dags att pröva signifikansen hos de samband vi tycker oss kunna se! Låt oss ta dummyvariabeln medellåg utbildning, UTB(1), som exempel. Som framgår av ovanstående formel behöver vi två värden för att kunna genomföra ett Wald-test: B-värdet samt standardfelet (standard error, S.E.) för den oberoende variabeln. Petar vi in värdena från tabellen i formeln får vi följande resultat:

$$\begin{aligned}\text{Wald} &= (\beta / \text{ASE})^2 \\ &= (B / \text{S.E.})^2 \\ &= (0,311 / 0,163)^2 \\ &= 3,640\end{aligned}$$

Här har vi alltså resultatet av vårt alldeles egna Wald-test. Jämför vi med vad SPSS kommer fram till i kolumn 4 ser vi en viss avvikelse (3,642). Förklaringen har med avrundningar att göra – eftersom SPSS avrundar B-värdena till tre decimaler så tillåter det oss inte att bli lika exakta som programmet själv.

⁷ Enkelriktade (på eng. "directional" eller "one-tailed") hypotestester innebär att man testar huruvida värdet på den testade β -koefficienten är *antingen* större eller mindre än 0 – m a o om vi visa om det rör sig om antingen ett negativt eller ett positivt samband. Vid det mer konservativa dubbelriktade hypotestestet ("nondirectional"; "two-tailed") är ambitionen endast att kunna uttala sig om huruvida den oberoende variabeln överhuvudtaget skiljer sig från 0, oavsett om den är större eller mindre (se Sirkin 2006:210f). För mer om hypotesprövningar, se Aneshensel (2002).

⁸ En kritik mot ASE är att de anses vara överoptimistiska. Aldrich & Nelson (1984:55) rekommenderar att man i stället för z-fördelningen använder Student's t-fördelningen, vilken genom att den har högre kritiska värden ger mer konservativa hypotestester. t-testet har samma funktion som z-testet men används när populationens standardavvikelse är okänd (se t.ex. Sirkin 2006, kap. 8) . Tuft (2000:35) avfärdar argumentet som märkligt eftersom det alltså förutsätter att man arbetar med så stora urval att z- respektive t-fördelningen blir ungefär lika stora.

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a			40,006	3	,000	
utb						
utb(1)	,311	,163	3,642	1	,056	1,364
utb(2)	,760	,175	18,771	1	,000	2,139
utb(3)	,952	,171	31,060	1	,000	2,592
Kön	,544	,112	23,705	1	,000	1,722
F106a	,170	,027	39,196	1	,000	1,185
Constant	-2,550	,228	125,236	1	,000	,078

a. Variable(s) entered on step 1: F106a.

Precis som i vanlig OLS-regression anges signifikansnivåerna automatiskt av SPSS (kolumn 6, "Sig"). Värdet för utb(1) är 0,056. Vi klarar alltså inte att uppfylla den 5-procentiga signifikansnivån (Resultatet från Wald-testet – 3,642 – är ju också mindre än det kritiska värdet, 3,84!). Det går följaktligen inte med denna säkerhetsnivå att särskilja personer med medellåg från referensgruppen (se ovan). Dessbättre, för oss och våra hypoteser, är övriga variabler i högsta grad signifikanta.

Resultatet av Wald-testet ges alltså direkt av SPSS. Vill vi istället göra ett z-test, vilket SPSS inte redovisar, blir resultatet så här:

$$z = \beta / \text{ASE} = 0,311 / 0,163 = 1,908$$

En alternativ väg är förstås att ta roten av resultatet från Wald-testet (Wald är som redan nämnts det samma som z^2). Vårt resultat, 1,908, är mindre än det kritiska värdet 1,96 och nollhypotesen kan inte heller här förkastas.

Innan vi bekantar oss med *ytterligare* ett alternativ sammanfattar vi:

- Wald-testet (och dess släkting z-testet) används i logistisk regression för att testa signifikansen hos *enskilda* koefficienter
- Wald är den kvadrerade kvoten av den ostandardiserade logistiska koefficienten, B, genom dess standardfel, S.E.
- Wald-testet motsvarar signifikanstestet av B-koefficienter i OLS-regression

3.3 Likelihood ratio test

Det vanligaste alternativet till z- och Wald-testen är likelihood ratio testet. Med likelihood ratio testet kan utföras åtminstone tre olika typer av signifikanstester:

1. Av bidraget av en *ny* variabel till en modell
2. Av *hela* modeller
3. Av *individuella* variabler

Wald kan som redan framgått endast användas för punkt 3, att studera signifikans hos enskilda koefficienter.

Principen bakom Log likelihood och Likelihood ratio

En "likelihood" uttrycker som namnet antyder en sannolikhet, i det här fallet den sannolikhet med vilken värdet på den beroende variabeln kan prediceras utifrån ett observerat värde på en oberoende variabel. På samma sätt som andra sannolikheter varierar likelihood mellan 0 och 1. Log likelihood (LL) är dess logaritmerade värde och varierar mellan 0 och minus oändligheten (det är negativt för att den naturliga logaritmen av alla tal under 1 är negativa). Eftersom $-2LL$ (alltså LL gånger -2) har en approximativ χ^2 -fördelning kan $-2LL$ användas för att testa signifikans i logistisk regression. Principen är här den samma som med Sum of squared errors (SSE) i OLS-regression. $-2LL$ kallas också Likelihood ratio.

Log likelihood beräknas med hjälp av den Maximum likelihood funktion på vilken den logistiska regressionen är baserad. Beräkningsmetoden går som redan nämnts ut på att hitta de koefficienter som maximerar logaritmen av Maximum likelihood-funktionen, dvs. log likelihood. Detta sker genom en rad iterationer ("upprepningar"). SPSS letar sig genom upprepade beräkningar fram till de koefficienter som maximerar log likelihood för den fullständiga modellen, dvs. sannolikheten för att man utifrån modellens en eller flera oberoende variabler ska kunna predicera värdet på den beroende variabeln. I denna process minskar värdet på $-2LL$ allteftersom modellen förbättras.

I SPSS visas $-2LL$ i kolumnen " -2 Log likelihood" i output-rutan "Model Summary". (Vill man även se samtliga iterationer måste man kryssa i rutan "Iteration history" under "Options" i kommandofönstret för logistisk regression.)

Ett likelihood ratio test görs genom att man testat om differensen mellan två likelihood ratios är signifikant eller inte.

Vi börjar med den enklaste av provningar: nämligen ett test av en modell som har en variabel jämfört med en modell som endast har en konstant. Eller annorlunda uttryckt: en provning av huruvida effekten av en oberoende variabel i en vanlig enkel bivariat logistisk regression är signifikant eller inte.

Som vanligt kan vi formulera två hypoteser (vi använder här vår dummykodade utbildningsvariabel som ensam oberoende variabel):

H_0 : Det finns inga skillnader mellan personer med olika utbildningsnivå när det gäller förtroende för svenska politiker

H_1 : Det finns en skillnad mellan olika utbildningsnivåer i fråga om politikerförtroende

Om vi kallar likelihood ration i ursprungsläget (dvs. en modell med endast en konstant) för $-2LL_0$ och likelihood ration när vi har inkluderat en variabel för $-2LL_1$, kan vi ställa upp likelihood ratio testet enligt följande formel:

$$G^2 = (-2LL_0) - (-2LL_1)$$

Vi får på så sätt en testoperator som kallas G^2 , vilket är ett χ^2 -värde. Likelihood ratio testet är som vi minns en typ av χ^2 -test, vilket är mycket vanligt inom statistiken (se t.ex. Sirkin 2006, kapitel 12).

Från SPSS hämtar vi nedanstående output-rutor. (Observera att vi gjort en körning där våra tre oberoende variabler placerats i olika *block*, dvs. Next-funktionen i kommandofönstret). Det ska även nämnas att likelihood-ratio testet presenteras automatiskt av programmet, inga särskilda ”ibockningar” behövs alltså. (För att visa själva iterationsprocessen bakom testet har vi dock i exemplet som följer kryssat i Iteration history under Options i fönstret för logistisk regression.)

En vanlig enkel bivariat logistisk regression redovisas i SPSS i två steg, eller block. I det första – nedan kallat Block 0 – ingår endast konstanten, m a o det som ovan kallas $-2LL_0$, och i steg två – Block 1 – visas vad som sker när vår oberoende variabel, $-2LL_1$, inkluderas.

Under rubriken Block 0 redovisas resultaten för ursprungsmodellen, alltså den i vilken endast konstanten ingår. Direkt under rubriken ser vi en ruta som heter Iteration history (den visas därför att vi tidigare valde att kryssa för denna ”tjänst”) . Vi ska här koncentrera oss på mittenkolumnen: -2 Log likelihood, med andra ord likelihood ratio. Här ser vi hur iterationsprocessen förlöpt: den sista iterationen (step 3), dvs. den som alltså maximerar log likelihood, ger oss vårt $-2LL_0$ -värde: 2012,410.

Block 0: Beginning Block

Iteration History^{a,b,c}

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	2012,585	-,631
	2	2012,410	-,653
	3	2012,410	-,653

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 2012,410

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Under nästa rubrik – Block 1 – har vår beroende variabel inkluderats. Från rutan Iteration history plockar vi vårt $-2LL_1$: 1960,485. Vi kan nu räkna fram vårt G^2 enligt den ovanstående formeln:

$$G^2 = (-2LL_0) - (-2LL_1) = 2012,410 - 1960,485 = 51,925$$

Här har vi alltså den förändring som vår utbildningsvariabel innebär för vår beroende variabel. Som vi ser medför införandet av variabeln utbildningsnivå en lägre likelihood ratio, vilket är bra: ambitionen är ju att minimera den. Frågan är då: är minskningen signifikant? SPSS har svaret. I den näst sista tabellen (Omnibus Tests of Model Coefficients) presenteras tre χ^2 -tester – Step, Block och Model. Som vi ser är de identiska. Värdet är 51,925 – som synes vårt egenhändigt framräknade G^2 ! – och visar sig vara signifikant med tre frihetsgrader (Antalet frihetsgrader är det samma som antalet oberoende variabler i modellen; eftersom vår utbildningsvariabel består av tre dummyvariabler får vi alltså tre frihetsgrader.). Gränsvärdena

för signifikanstestet ($p \leq$) går vid 0,05, 0,01 eller 0,001 beroende på vilken säkerhetsmarginal man eftersträvar. Inget nytt här alltså!

Block 1: Method = Enter

Iteration History^{a,b,c,d}

Iteration		-2 Log likelihood	Coefficients			
			Constant	utb(1)	utb(2)	utb(3)
Step 1	1	1961,740	-1,072	,264	,638	,936
	2	1960,486	-1,194	,337	,753	1,057
	3	1960,485	-1,198	,340	,756	1,061
	4	1960,485	-1,198	,340	,756	1,061

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 2012,410

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	51,925	3	,000
	Block	51,925	3	,000
	Model	51,925	3	,000

Vårt test visar följaktligen att genom att inkludera utbildningsnivå får vi en statistiskt signifikant förbättring av likelihood ratio jämfört med en modell som bara innehåller konstanten. Vi kan alltså förkasta vår nollhypotes om att det inte finns något samband mellan utbildningsnivå och politikerförtroende. För det gör det.

Likelihood ratio testet ger också möjligheten att pröva om ytterligare utökningar av modellen innebär någon signifikant förbättring av likelihood-funktionen. Vad händer till exempel med sambandet mellan utbildningsnivå och politikerförtroende om man även tar hänsyn till kön? Vi testar följande hypoteser:

H_0 : Kön har ingen betydelse för sambandet mellan utbildningsnivå och politikerförtroende. (Information om kön bidrar inte till en bättre predicering av utfallet på den beroende variabeln utöver informationen om utbildningsnivå.)

H_1 : Kön har betydelse för sambandet mellan utbildningsnivå och politikerförtroende.

Här nedan följer ett utdrag ur SPSS för vår trivariata modell. Observera att vi för att kunna undersöka denna förändring i SPSS måste välja att göra analysen i två steg, eller block (klicka på Next i regressionsfönstret och peta sedan in könsvariabeln i den tomma högerkolumnen). Därmed får vi utöver Block 0 (endast konstanten) och Block 1 (konstant + 1 oberoende variabel) också ett Block 2 (konstant + 2 oberoende variabler). Då de två första blocken är identiska med vad som tidigare presenterats koncentrerar vi oss på Block 2.

Block 2: Method = Enter

Iteration History ^{a,b,c,d}							
Iteration		-2 Log likelihood	Coefficients				
			Constant	utb(1)	utb(2)	utb(3)	Kön_NY
Step 1	1	1938,272	-1,327	,292	,696	,983	,466
	2	1936,207	-1,504	,373	,829	1,127	,537
	3	1936,204	-1,511	,377	,835	1,133	,539
	4	1936,204	-1,511	,377	,835	1,133	,539

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 1960,485

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Vi börjar återigen med att kika på Iteration History. Vi har här fått ett nytt -2LL värde: 1938,272. Det är likelihood ration för hela den trivariata modellen, med andra ord: -2LL₂. Vi kan nu beräkna den eventuella förändring (uttryckt i χ^2 , χ^2) som könsvariabeln medför:

$$\chi^2 = (-2LL_1) - (-2LL_2) = 1960,485 - 1936,204 = 24,281$$

Nu behövde vi egentligen inte räkna ut detta själva, svaret ges av SPSS i den nästföljande tabellen (förutsatt förstås att vi använt oss av ovan nämnda Block-funktion!). Också här blir resultatet ett lägre χ^2 – det tycks alltså finnas en effekt även av variabeln kön.

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	24,281	1	,000
	Block	24,281	1	,000
	Model	76,205	4	,000

Nu ser vi att de olika χ^2 -värdena i tabellen plötsligt skiljer sig åt. I raden Block återfinner vi vårt redan beräknade tillskott av könsvariabeln. Som synes är förändringen signifikant (antalet frihetsgrader utgörs av antalet nya variabler, alltså 1). I den sista raden signifikantestast så

hela modellen.⁹ Värdet, 76,205, utgörs helt enkelt av summan av förändringen av utbildnings- och könsvariablerna (51,925 + 24,281). Också här är förändringen signifikant (3 + 1 frihetsgrader = 4).

Vi avslutar med att plocka in vår sista förklaringsvariabel, mellanmänsklig tillit. Också denna visar sig kunna bidra med en förändring ($\chi^2 = 41,915$) som är statistiskt signifikant:

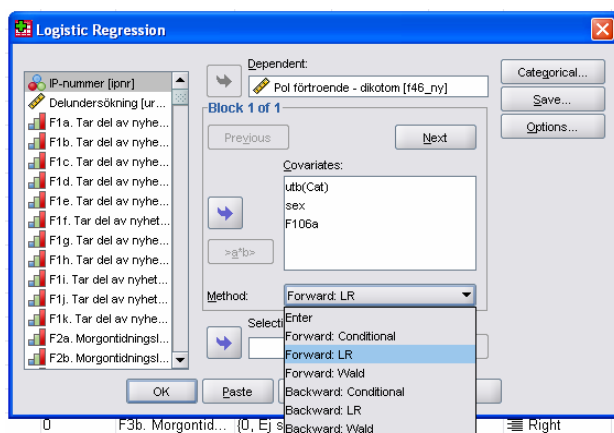
Block 3: Method = Enter

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	41,915	1	,000
	Block	41,915	1	,000
	Model	118,120	5	,000

Observera: Chi2-testet Model i figuren ovan ska endast ses som ett test av hur modellens koefficienter *gemensamt* kan bidra till att förutspå värdet på den beroende variabeln. Med det menas att ett signifikant värde (som i illustrationen här ovan) endast kan tolkas som att det finns ett signifikant samband mellan *åtminstone en* av modellens oberoende variabler och den beroende variabeln. För att säkerställa att samtliga oberoende variabler faktiskt är signifikanta måste man ta till blockfunktionen, på samma sätt som vi gjort det här presenterade exemplet. Ett alternativ till denna funktion är att använda funktionen Step i SPSS (Stepwise logistic regression), som fungerar på ungefär samma sätt som block (se Garson).

Vi har kunnat se att likelihood ratio testet kan användas för att testa signifikansen hos dels hela modeller, dels skillnaden mellan modeller (i bemärkelsen bidraget av en ny variabel som tillförs en modell). Avslutningsvis kan vi också använda testet för att pröva huruvida vissa oberoende variabler är viktigare än andra. Det går till så att man testat vad som händer om man *tar bort* en variabel från modellen.

För detta krävs att vi använder SPSS step-funktion. Under Method (se figuren) väljer vi att klicka för Forward:LR. LR står här för Likelihood ratio.



⁹ Figurens första rad, "Step", behöver man bara bry sig om ifall man använder Step-funktionen i SPSS (se nedan). Så länge man inte gör det visar den alltid samma resultat som "Block".

I tabellen nedan (Variables in the Equation) ser vi hur SPSS genom den stegvisa funktionen testat tre modeller (Step 1-3; f106a är variabeln för mellanmänsklig tillit); för varje steg adderas en oberoende variabel. Turordningen baseras på det förvalda kriteriet bäst Likelihood Ratio.

Variables in the Equation		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	F106a	,193	,026	53,489	1	,000	1,213
	Constant	-1,925	,186	106,705	1	,000	,146
Step 2 ^b	utb			35,063	3	,000	
	utb(1)	,275	,161	2,893	1	,089	1,316
	utb(2)	,682	,173	15,539	1	,000	1,979
	utb(3)	,880	,169	27,170	1	,000	2,411
	F106a	,169	,027	39,444	1	,000	1,184
	Constant	-2,226	,214	108,503	1	,000	,108
Step 3 ^c	utb			40,006	3	,000	
	utb(1)	,311	,163	3,642	1	,056	1,364
	utb(2)	,760	,175	18,771	1	,000	2,139
	utb(3)	,952	,171	31,060	1	,000	2,592
	Kön	,544	,112	23,705	1	,000	1,722
	F106a	,170	,027	39,196	1	,000	1,185
	Constant	-2,550	,228	125,236	1	,000	,078

a. Variable(s) entered on step 1: F106a.

b. Variable(s) entered on step 2: utb.

c. Variable(s) entered on step 3: Kön.

Vi riktar nu vår uppmärksamhet till nästa tabell: Model if Term Removed (nedan). För varje rad provas vad som händer med de olika modellerna om en av variablerna exkluderas. En förändring som *inte* är signifikant (se sista kolumnen) indikerar att det inte finns någon skillnad mellan den fullständiga och den reducerade modellen. Variabeln kan därför tas bort och man får då en mindre komplicerad modell utan att göra avkall på förklaringskraften.

Model if Term Removed				
Variable	Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
Step 1 F106a	-1006,205	58,454	1	,000
Step 2 utb	-976,978	35,675	3	,000
F106a	-980,243	42,204	1	,000
Step 3 utb	-967,622	40,954	3	,000
Kön	-959,140	23,991	1	,000
F106a	-968,102	41,915	1	,000

Som framgår av tabellen här ovan får vi i samtliga tre modeller signifikanta värden för alla koefficienter. Detta säger oss att alla våra tre variabler har en signifikant förklaringskraft och bör behållas i modellen.

Men inte nog med det. I kolumnen Change in -2 Log Likelihood anges χ^2 -värden för hur stor förklaringskraft som går förlorad om variabeln i fråga tas bort från modellen. Ser vi till hela vår modell (Step 3) kan vi alltså konstatera att mellanmänsklig tillit (41,915) är den variabel som bäst bidrar till att förutspå utfallet av vår beroende variabel. Den följs tätt av utbildning. Kön är som synes den variabel som tillför minst till vår modell, även om förklaringskraften också här är signifikant.

Något om forward respektive backward stepwise regression

SPSS kan analysera datamaterialet genom stegvisa modeller, s.k. stepwise regression. Valet görs i rutan "Method" i fönstret för logistisk regression (se figur ovan). *Framåtriktade* analyser ("Forward") är vanligast. Utgångspunkten är en modell med endast konstanten. Utifrån ett särskilt kriterium (till exempel bästa Likelihood Ratio, Wald etc.) adderar sedan programmet en oberoende variabel i sänder ända tills en särskild nivå är nådd (t.ex. det steg vid vilket samtliga variabler som inte ingår i modellen har en signifikansnivå som är högre än 0,05; nivån sätts under "Probability for "Stepwise" under "Options"; görs inget aktivt val används 0,05-nivån). *Bakåtriktade* analyser ("Backward") startar med samtliga variabler och exkluderar sedan en i taget, i den ordning som de passar sämst enligt det valda kriteriet.

Observera: Stegvisa analyser är lockande i så måtto att den inbjuder till jakt efter den "perfekta" modellen, modellen med största möjliga förklaringskraft. Man kan ju stoppa in alla tänkbara variabler och sedan se vad som kommer ut! Vår rekommendation är likafullt att man bör vara restriktiv med funktionen och endast använda den i explorativa syften. Den seriöse och allvarligt menande samhällsforskaren arbetar givetvis med teoristyrda modeller och kör således företrädesvis sina logistiska regressioner med den förinställda Enter-funktionen.

Innan diskussionen om vilket test som är att föredra sammanfattar vi:

- Likelihood ratio testet är ett signifikanstest som fungerar enligt χ^2 -principen.
- I likhet med Wald-testet kan det användas för att testa signifikansen hos enskilda koefficienter, men också (och till skillnad från Wald), signifikansen hos hela modeller och mellan olika modeller.
- Likelihood ratio testet motsvarar F-testet i OLS-regression.

3.4 Wald eller Likelihood ratio – vad är bäst?

Så kommer vi då till den oundvikliga frågan: Wald eller Likelihood ratio – vad är att föredra? Som vi redan varit inne på fungerar Wald-testet endast på enskilda koefficienter, medan Likelihood ratio testet även klarar att testa hela modeller. Den senare har alltså ett bredare tillämpningsområde, vilket i sig är en viktig fördel.

För att använda den senare talar två saker: Agresti (1996) skriver: “For very large samples, [the two tests] have similar behaviour. For sample sizes used in practice, the likelihood-ratio test is usually more reliable than the Wald test” (89). Det finns alltså skäl att vara försiktig med z- och Wald-test när man har att göra med mindre urval.

Men vad är då ett stort alternativt litet urval? Litteraturen, konstaterar Tufte, ger dessvärre ”få rettesnorer” (2000:36). Long (1997:53) avråder från att göra logistisk regression på urval med färre än 100 observationer, och att urval på över 500 observationer får anses betryggande. Long menar dock att det är modellens *kovariatstruktur* som avgör hur många observationer som krävs. Kovariatstrukturen kan enkelt beskrivas som analysmodellens komplexitet, dvs. antalet oberoende variabler och antalet variabelvärden hos dessa variabler (som vi redan noterat kallas variablerna i logistisk regression för covariates). Ju fler variabler och variabelvärden, ju mer komplex kovariatstruktur, och ju komplexare kovariatstruktur, desto högre urvalskrav. Long delar här med sig av en tumregel: för varje parameter i modellen, dvs. för varje variabel eller konstant, bör det vara åtminstone tio observationer. Härutöver kan det vara nödvändigt med ännu fler observationer då det är liten *varians* i den beroende variabeln eller då finns en tydlig *kollinearitet*. På samma sätt fordrar vissa *typer* av regressioner, t.ex. då den beroende variabeln befinner sig på ordinalskalenivå, också fler observationer. Sammanfattningsvis så finns det alltså flera – om än nog så oprecisa – antydningar om hur stort urval som krävs för att den logistiska regressionen ska vara meningsfull som metod.

Hosmer & Lemeshow (1989:17), liksom Menard (1995:39, i Garson 2009) påpekar en annan svaghet med z- och Wald-testen, nämligen att deras pålitlighet sjunker i analyser då det *absoluta värdet på den erhållna logit koefficienten är mycket högt*. De hänvisar till studier som visar att de båda testen inte sällan ger missvisande bedömningar när det gäller huruvida nollhypotesen kan förkastas. Med andra ord tenderar de att bekräfta nollhypotesen om inget samband, när det i själva verket är tvärtom. Anledningen är enligt författarna att höga absoluta värden på koefficienten ger alltför höga standardfel, vilket i förlängningen innebär att det blir svårare att klara de kritiska värdena för testen. Risker ökar därmed för s.k. typ II fel (dvs. att felaktigt avfärda en effekt som ej signifikant). Enligt Garson blir en konsekvens av detta att om man använder sig av dummyvariabler är det säkrare att testa skillnaden mellan en modell där dummyvariabeln är med och en där den exkluderats.

Så samtliga här redovisade statistiker tycks alltså förorda likelihood ratio testet framför Wald-testet när det gäller tester av enskilda koefficienter. Samtidigt konstaterar vi att Wald-testet är nog så ofta tillämpat i litteraturen.

Avslutningsvis ska man komma ihåg att såväl likelihood ratio som Wald är s.k. asymptotiskt ekvivalenta, med vilket menas att de alla konvergerar mot samma χ^2 -fördelning alltefter att antalet observationer ökar. Vi ska inte fördjupa oss i vad detta innebär, utan nöjer oss med att konstatera att logistiska regressioner – oavsett signifikanstestmetod – kräver relativt stora urval.

3.5 Hur "bra" är modellen? Hosmer och Lemeshow ger svar.

Eftersom en bra modell överensstämmer med verkligheten ställs vi inför ett dilemma: vår modell har skapats eftersom vi *inte* vet hur verkligheten ser ut, vad vi utgår från är *en del* av verkligheten. En pragmatisk lösning är emellertid att jämföra predicerade värden med de värden som vi faktiskt har, alltså de data som ligger till grund för modellen.

Hosmer & Lemeshow (1989) har lanserat en metod där predicerade värden *gruppvis* jämförs med observerade värden.¹⁰ Testet kan enkelt beställas i SPSS – allt du behöver göra är att under Options välja Hosmer-Lemeshow Goodness of Fit Test.

Nedan kan vi se resultatet av ett Hosmer och Lemeshow-test för den modell som vi har prövat.

Contingency Table for Hosmer and Lemeshow Test

		Pol förtroende = 0		Pol förtroende = 1		Total
		Observed	Expected	Observed	Expected	
Step 1	1	141	141,694	24	23,306	165
	2	120	126,396	39	32,604	159
	3	118	115,438	35	37,562	153
	4	107	112,496	50	44,504	157
	5	114	109,510	46	50,490	160
	6	121	116,288	61	65,712	182
	7	102	104,455	73	70,545	175
	8	103	92,050	65	75,950	168
	9	73	76,753	87	83,247	160
	10	31	34,919	56	52,081	87

Vad boxen visar är att

- Tio grupper har urskiljts
- Grupperna är snarlika i storlek¹¹
- Ingenstans stämmer de observerade och de predicerade värdena perfekt, modellens förutsägelser skiljer sig från hur verkligheten ser ut.

Frågan som uppstår är hur stor denna skillnad egentligen är. Är skillnaden mellan predicerade och observerade värden för stor kommer vår modell att underkännas, är skillnaden på en acceptabel nivå kommer vår modell att ha klarat testet. Jämfört med provandet av en

¹⁰ Principen bakom gruppindelningen är rangordning. Utifrån predicerade sannolikheter urskiljs 10 grupper; där sannolikheten för $y=1$ är stor kommer sannolikheten för $y=0$ att vara liten och vice versa (se tabell). För varje grupp jämförs sedan det antal som *prediceras* ha högt (respektive lågt) förtroende med det antal som *observeras* ha högt (respektive lågt) förtroende. I en modell med perfekt passning finns det ingen skillnad mellan förväntade och observerade antal, i en modell med dålig passning är skillnaderna genomgående stora.

¹¹ Att grupperna inte är identiska i storlek förklaras av att vår beroende variabel har få värden (jfr Tufte 2001:47). Varför grupp 10 avviker med endast 87 fall har vi – tyvärr! – inget bra svar på.

nollhypotes är principen alltså den omvända: en signifikansnivå *över* 0,05 innebär "ett godkännande", skillnaden är inte större än vad som kan förklaras av slumpen.

Låt oss återvända till vårt exempel. De oberoende variablerna är, precis som tidigare, utbildning (dummykodad), kön och mellanmänsklig tillit.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	8,005	8	,433

Tabellen visar resultatet av χ^2 -testet (8,005), antalet frihetsgrader (8) och den signifikansnivå som detta motsvarar (0,433). 0,433 är *större* än 0,05; vad testet visar är alltså att modellen som helhet ger värden som inte avviker mer från de faktiska än vad som skulle kan förklaras av slumpen.

Hur χ^2 -värdet beräknas? Den som vill kan själv göra beräkningen för hand!

1. Cell för cell jämförs observerade värden med förväntade värden. I Contingency-tabellen ovan är det observerade värdet (antalet) i cellen längst upp till vänster 141, medan det förväntade värdet (antalet) är 141,694. Differensen mellan det observerade och det förväntade värdet blir således $141 - 141,694 = -0,694$.
2. I nästa steg kvadreras den erhållna differensen ($-0,694^2 = 0,482$).
3. Den kvadrerade differensen divideras med det förväntade värdet ($0,482/141,694 \approx 0,003$).
4. Proceduren upprepas för samtliga celler (= 20 stycken), varpå de erhållna värdena adderas. Summan utgör vårt χ^2 -värde.

Varför ett värde över 0,05 innebär ett godkännande!

Att ett värde över 0,05 innebär att modellen godkänns kan kännas lite konstigt – vanligtvis förknippar vi ju höga signifikansvärden som en sorts avslag: "Nej, tyvärr, gå tillbaks och tänk om! Variabeln saknar effekt på det som ska förklaras!"

Förklaringen är att själva grundtänket är det omvända: För att förkasta en nollhypotes krävs det att effekten är större än vad som kan förklaras av slumpen ($p < 0,05$). För att en modell ska accepteras krävs det att skillnaden mellan observerade och predicerade värden **inte** är större än vad som kan förklaras av slumpen ($p > 0,05$).

Att en modell "godkänns" innebär inte att den nödvändigtvis förklarar en stor del av variansen i vår beroende variabel, bara att andelen – hur stor eller liten den nu må vara! – är signifikant.

3.6 Relevansen av R^2

Ovan har det framgått att logistisk regression främst är en användbar metod för prövningen av hypoteser. Att hypotesprövningen är i fokus innebär att logistisk regression kan uppfattas som en variabel-/faktororienterad metod. Med denna utgångspunkt följer att andelen förklarad varians är av underordnad betydelse. Det intressanta är inte att förklara så mycket som möjligt av utfallet – vad som intresserar oss är att studera effekten av enskilda (teoretiskt förankrade) variabler (jfr Hagquist & Stenbeck 1998).

Mot bakgrund av detta är det inte särskilt konstigt att det saknas tydliga normer kring vilket mått som bör användas för att bedöma hela modeller (se faktarutan nedan). SPSS redovisar två mått: Cox och Snells R^2 samt Nagelkerkes R^2 . Testen presenteras i anslutning till Hosmer och Lemeshow-testet, hur outputen ser ut framgår nedan.¹²

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1894,290 ^a	,073	,100

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

I princip bygger både Nagelkerkes och Cox och Snells metod på att den slutliga modellen jämförs med ursprungsmodellen, d.v.s. den modell som utgörs av enbart konstanten¹³. I valet mellan att ange Cox och Snells eller Nagelkerkes R^2 föredrar de flesta att använda Nagelkerkes mått, ”det mått som är mest analogt med determinationskoefficienten R^2 i vanlig OLS-regression” (Djurfelt & Barmark 2009:132, n 6).

Men *går* det då att tolka de erhållna värdena som andelen förklarad varians? Också här är buden lite olika. I kommentarerna till ett Nagelkerke-värde på 0,107 skriver Djurfeldt & Barmark (2009:137) att ”10,7 % av den totala informationen kan föras tillbaka på [...] de förklarande variablerna” – tämligen oproblematiskt förefaller alltså Nagelkerke-värdet ha omvandlats till procentsatser.

Garson intar motsatt ståndpunkt och skriver att pseudomåtten ”should be reported as approximations to OLS R^2 , not as actual percent of variance explained”. Även om han inte syftar på just Nagelkerke-måttet (rekommendationen gäller alla pseudo R^2) är han alltså avgjort kritisk till att omvandla pseudo R^2 till procentandelar.

¹² Vare sig Cox och Snells eller Nagelkerkes R^2 räknas som ”riktiga” R^2 , i Stanotes kallas de för ”pseudo R^2 ” och Tuft (2000:47) förespråkar istället ett sätt som bygger på en bivariat regression mellan observerade värden och predicerade sannolikheter. Eftersom meningsfullheten i R^2 kan ifrågasättas väljer vi emellertid att inte gå in på metoden (som f.ö. kräver en viss manuell bearbetning av data).

¹³ För den som tycker att vi uttrycker oss alltför vagt väljer vi att hänvisa till en sida från UCLA: http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Pseudo_RSquareds.htm Utöver Nagelkerkes samt Cox och Snells (pseudo) R^2 presenteras här en rad andra mått, t.ex. de som utarbetats av McKelvey och Zavoina, McFadden och Efron. Trevlig läsning!

”Betydelsen av den beroende variabelns fördelning”

Att det inte finns något allmänt accepterat mått som är analogt till OLS-regressionens R^2 förklaras av att variansen i en dikotom beroende variabel är avhängig hur fördelningen ser ut. Ju mer snedfördelade variablerna är, desto mindre blir den varians som ska förklaras.

Detta innebär givetvis att användbarheten av så kallade ”pseudo R^2 ” är begränsad, måtten kan till exempel inte användas för att jämföra regressioner där de beroende variablerna har olika fördelning.

Innan vi går vidare med hur resultatredovisningen kan se ut sammanfattar vi:

- Hosmer och Lemeshow-testet jämför predicerade värden med observerade värden. Ett χ^2 -värde som motsvarar en signifikansnivå över 0,05 innebär att modellen är ”bättre” än slumpen.
- Genom att Hosmer och Lemeshow-testet bygger på grupperade data fungerar ”Contingency-tabellen” som en indikator på passningsgraden för olika ”intervall”.
- Det pseudo R^2 som vanligtvis används är Nagelkerkes. Huruvida måttet verkligen kan omvandlas till procentandelar är omdiskuterat.

4.1 Resultaten presenteras

Resultat från logistiska regressionsanalyser kan på samma sätt som resultat från OLS-regressioner presenteras i form av kurvor. Förutsatt att man inte har alltför många förklaringsvariabler är detta det absolut mest lättillgängliga sättet för grafisk framställning av logistiska regressionsanalyser.

Så här går det till. Först och främst behövs ett räkneprogram, vi har i det följande använt oss av Excel. I figuren här nedanför har vi använt resultaten från vår modell för att förklara förtroendet för svenska politiker. För att slippa tredimensionella figurer har vi valt att redovisa effekten av utbildning och mellanmänsklig tillit på vår beroende variabel i två separata figurer, en för kvinnor, och en för män.

Observera: Det finns en avgörande skillnad mellan hur resultaten från OLS respektive logistisk regression presenteras, och det är det *mått* som används för att uttrycka den beroende variabeln. Som bekant visar OLS förändringen i den beroende variabeln av ett skalstegs förändring av en oberoende variabel. Med andra ord används det mått med vilket den beroende variabeln mäts, till exempel vikt i kilogram: En burhöna som levt i fem månader och matats med tre skopor säd om dagen kan väntas väga 3,7 kg. Vid logistisk regression består istället måttet av *sannolikheten* att något ska inträffa, exempelvis: Sannolikheten att en man med hög utbildning ska prenumerera på en morgontidning uppgår till 0,7. Som andra sannolikhetsmått rör det sig om en skala mellan 0 och 1 (givetvis funkar det också att använda sig av procentsatser). En annan minst lika viktig skillnad är att det i logistisk regression inte finns några krav på *linjäritet*: Vi kan utan problem få ”kurviga” kurvor.

Figur 4.1: Tillvägagångssättet för beräkning av sannolikheter. Utdrag ur Excel.

	A	B	C	D	E	F	G
1	Variabel	Koefficienter					
2							
3	Konstant	-2,550					
4	Utb(1)	0,311					
5	Utb(2)	0,760					
6	Utb(3)	0,952					
7	Sex	0,544					
8	f106a	0,170					
9							
10		Kvinnor					
11		Låg utb.		Medellåg utb.		Medelhög utb.	
12	Mellanm. tillit	Logit	Sannolikhet	Logit	Sannolikhet	Logit	Sannolikhet
13	0	-2,550	0,07241211	-2,240	0,09625109	-1,790	0,14308
14	1	-2,380	0,08467572	-2,070	0,11206428	-1,620	0,165188
15	2	-2,211	0,09879507	-1,900	0,13010144	-1,450	0,18994
16	3	-2,041	0,11497305	-1,730	0,15054952	-1,281	0,217450
17	4	-1,871	0,13340805	-1,561	0,17367018	-1,111	0,247717

= B3

= 1 / (1+EKS(A13))

=\$B\$3 + \$B\$4 + (\$B\$8*A14)

Excelarket i figur 4.1 har strukturerats på samma sätt som om man skulle göra en vanlig korstabell. Arket har fyra beståndsdelar:

1. Modellens *variabler och koefficienter* (området A3:B8)
2. Värdet på våra *två oberoende variabler*: Utbildning (raden 11 från cell B och framåt) samt Mellanmänsklig tillit (kolumn A från cell 13 och nedåt)
3. Beräkningen av *logit* (kolumn B, D, F osv. från cell 13 och nedåt)
4. Beräkningen av *sannolikheter* (kolumn A, C, G osv. från cell 13 och nedåt)

I det första steget lägger vi in modellens *variabler och koefficienter* (dvs. log oddsen från kolumnen B). Dessa hämtas direkt ifrån outputfönstret i SPSS (se avsnitt 2.1). I nästa steg skriver vi in värdena på våra *två beroende variabler*; i lodrätt riktning placeras den elvagrada skalan för Mellanmänsklig tillit och i vågrätt riktning de fyra utbildningsnivåerna. (Notera att vi delat upp modellen i två delar, en för män och en för kvinnor. Exemplet baseras på resultatet för kvinnor. Sedan upprepas samma procedur för män.)

Nu är det dags att börja räkna! Först måste vi få fram *logiten* för varje kombination av värden för de två beroende variablerna. Det är dessa värden som senare utgör grunden för de sannolikheter som ska plottas in i våra två figurer. Utgångspunkten är formeln för analysmodellen, i vårt fall:

$$\text{Logit} = \text{Konstant} + \beta_{\text{utb}} + \beta_{\text{mmt}} + \beta_{\text{kön}}$$

Principen är precis densamma som för OLS-regression. Vi börjar från början med värdena 0 för Mellanmänsklig tillit samt utbildning Låg, vilka alltså ska in i cell B13 (se figur). Eftersom Låg utbildning är referenskategori i vår dummykodade utbildningsvariabel har den inget eget värde. Det samma gäller för Kvinna (Kvinna=0, Man=1). Samtliga tre variabler har

i det här fallet alltså värdet 0, vilket innebär att endast konstanten, -2,550, återstår. Det blir följaktligen den som vi skriver in i B13. Gör vi sedan kolumnen komplett så förblir såväl köns- som utbildningsvariabeln oförändrade (dvs. 0) medan värdet på tillitsvariabeln ökar till och med värdet 10.

Nästa exempel tar vi från rutan D14 (se figuren), där vi har värden för Utbildning (Medellåg=1) och Mellanmänsklig tillit (1), men inte för Kön (0). Det ger oss följande:

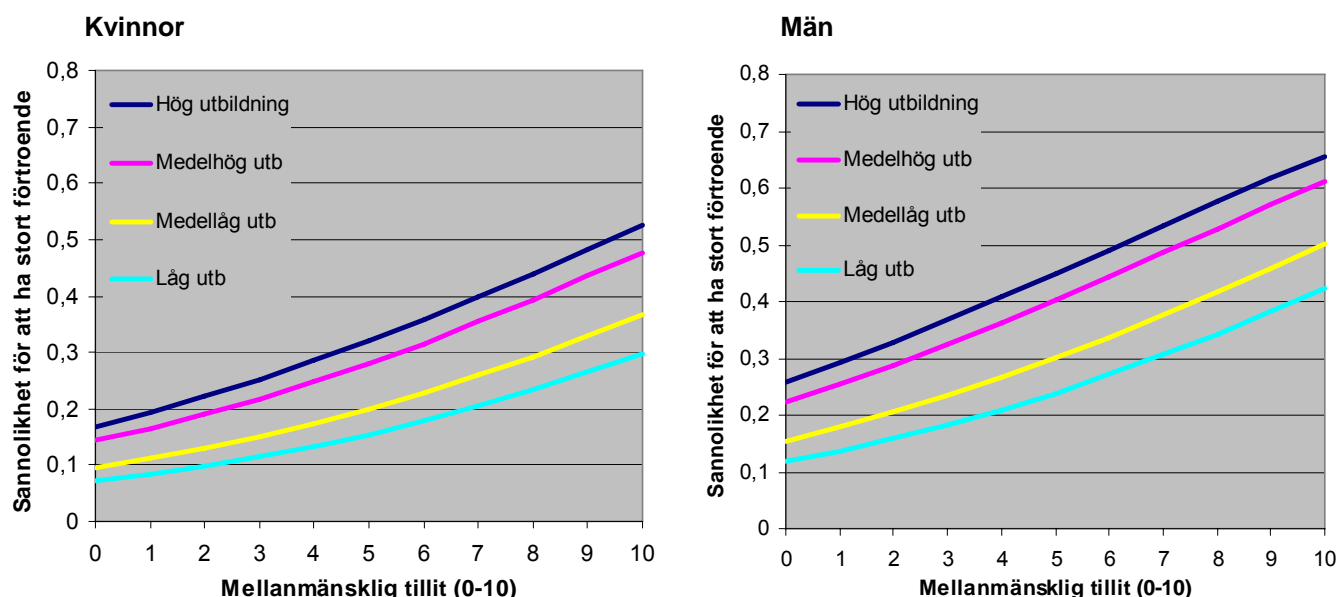
$$\text{Logit}_{D14} = -2,550 + (1 \times 0,311) + (0 \times 0,544) + (1 \times 0,170) = -2,070$$

Och på den vägen är det. Vi rekommenderar varmt att använda Excels formelkopieringsfunktion (se exempelformeln i den gröna ovalen; \$-tecknen i t.ex. "\$B\$3" innebär att när formeln kopieras in i en annan ruta så fortsätter värdet att hämtas från samma ställe, i det här fallet cellen B3, vilken innehåller värdet på vår konstant).

I det fjärde och avslutande steget använder vi våra logit-värden för att räkna fram *sannolikheten* för samtliga möjliga utfallskombinationer av de beroende variablerna. Som vi minns beräknas dessa med formeln $p = o/1+o$. Hur formeln skrivs i Excel är markerat i rött i figuren. Vi börjar nu närma oss måls ände: Med utgångspunkt i den rödmarkerade cellen C13 kan vi alltså notera att sannolikheten för att en kvinna med låg utbildning helt utan mellanmänsklig tillit (0 på den 11-gradiga skalan) ska ha stort förtroende för svenska politiker endast är 0,072 – dvs. omkring sju procent.

Talen i sannolikhetskolumnerna kan nu användas för att grafiskt presentera betydelsen av utbildning, kön och mellanmänsklig tillit på förtroende för svenska politiker. I och med att vi beräknat sannolikheterna för kvinnor och män var för sig får vi två parallella diagram:

Figur 4.2: Uppskattad sannolikhet (0-1) bland kvinnor respektive män att ha stort förtroende för svenska politiker, efter utbildningsnivå och grad av mellanmänsklig tillit (2007)



Kommentar: Data från den nationella SOM-undersökningen 2007. Den fyragradiga svarsskalan till frågan om förtroende för svenska politiker har dikotomiserats ("mycket litet" samt "ganska litet" förtroende = 0; "ganska" samt "mycket stort" förtroende=1). I figuren visas sannolikheten för att ha stort förtroende (1).

Nu ser vi tydligt effekten av våra tre förklaringsvariabler! Vi konstaterar att den uppskattade sannolikheten för att en person ska ha stort förtroende för politiker ökar med såväl grad av mellanmänsklig tillit som utbildningsnivå (notera hur skillnaden mellan de fyra utbildningsnivåerna ökar för varje steg på skalan för mellanmänsklig tillit). För samtliga nivåer är dessutom sannolikheten högre bland män än bland kvinnor. Skillnaderna är i vissa fall betydande: som vi minns är sannolikheten endast sju procent att en lågutbildad kvinna med låg mellanmänsklig tillit ska ha stort förtroende för våra svenska politiker. Motsvarande andel för en högutbildad man med hög tillit till andra människor är som vi ser i högerfigurens övre hörn nära 70 procent (66).

Slutligen konstaterar vi att det inte rör sig om några linjära samband: lutningen på samtliga kurvor ökar istället exponentiellt (eller kanske rättare: logaritmiskt). Den absoluta ökningen för varje enskilt skalsteg blir som synes hela tiden större. Dessa förhållanden hade aldrig kunnat klarläggas om vi istället förlitat oss på OLS-regressionen i vår analys.

4.2 Övrig redovisningsformalia

Garson konstaterar att det finns förvånansvärt lite konsensus angående exakt vilka mått och värden som ska/bör presenteras i vetenskapliga artiklar baserade på logistisk regressionsanalys. Istället lutar han sig i sin rekommendation till det förslag som ges av Peng, Lee & Ingersoll (2002). Förslaget avser tabeller och tycks följa devisen: Allt ska med! En något mindre uttömmande version (anpassad till de analysverktyg som vi gått igenom här och till vad som kan genomföras i SPSS) kan sammanfattas så här:

I tabell:

1. B-koefficient, standardfel (S.E.), Wald's χ^2 , frihetsgrader (df), signifikansnivån (p-värde) samt odds ratio ($\exp[b]$) för såväl konstanten som samtliga testade variabler. Med andra ord: *samtliga värden* som ges av SPSS i output-rutan Variables in the equation.
2. Tester av model-fit (t.ex. likelihood ratio, score, Hosmer & Lemeshow) med sina respektive χ^2 -resultat, signifikansnivåer (p) och frihetsgrader.

I fotnot:

1. Information om den beroende variabeln (hur den kodats, vilken som är referenskategori, etc.)
2. Pseudo- R^2 (Cox & Snell eller Nagelkerke)

Självklart glömmer vi inte heller n-värdet! Med utgångspunkt i de analyser som gjorts tidigare i framställningen kan en redovisning av en logistisk regressionsanalys se ut enligt följande:

Tabell 4.1 Logistisk regressionsanalys av förtroendet för svenska politiker, efter utbildningsnivå, grad av mellanmännisklig tillit och kön (2007)

Koefficient	β	Standardfel (S.E.)	Wald's χ^2	Frihetsgrader (df)	p	Odds ratio (e^β)
Konstant	-2,550	0,228	125,236	1	0,000	0,078
Utbildningsnivå (dummykodad; ref. = låg)			40,006	3	0,000	
Medellåg	0,311	0,163	3,642	1	0,056	1,364
Medelhög	0,760	0,175	18,771	1	0,000	2,139
Hög	0,952	0,171	31,060	1	0,000	2,592
Kön (0=kvinna, 1=man)	0,544	0,112	23,705	1	0,000	1,722
Mellanmännisklig tillit (0–10)	0,170	0,027	39,196	1	0,000	1,185
Test			χ^2	df	p	
Overall model evaluation						
Likelihood ratio test			118,120	5	0,000	
Goodness-of-fit test						
Hosmer & Lemeshow			8,005	8	0,433	

Kommentar: Data från den nationella SOM-undersökningen 2007. Den fyragradiga svarsskalan till frågan om förtroende för svenska politiker har dikotomiserats ("mycket litet" samt "ganska litet" förtroende = 0; "ganska" samt "mycket stort" förtroende=1). Cox & Snell's $R^2 = 0,073$. Nagelkerke's $R^2 = 0,100$. n = 1 566.

5.1 Sammanfattning och slutsatser

I inledningen presenterade vi tre anledningar till varför logistisk regression är en metod som lämpar sig synnerligen väl för samhällsvetenskaplig forskning.

För det första angavs att metoden kan hantera *kvalitativa data*.

För det andra framgick det att metoden kan hantera också *snedfördelade mängder*.

För det tredje lyfte vi fram att relationen mellan de oberoende variablerna och den beroende variabeln *inte* behöver vara linjär, metoden fungerar också på samband som är *kurvlinjära*.

Vi har valt att arbeta med ett exempel där den beroende variabeln är sannolikheten för stort politikerförtroende. Andra exempel som skulle ha varit fullt lika möjliga är t.ex. sannolikheten för tidningsprenumeration, deltagande i de allmänna valen, eftergymnasiala studier, ensamhushåll, regelbundna konsert- och/eller biobesök, körkortsinnehav, och så vidare...¹⁴

¹⁴ Att vi har valt att redogöra för den *binominala* logistiska regressionen är ett resultat av att många av samhällsvetenskapens variabler faktiskt är *dikotoma*: antingen så prenumererar man på en tidning eller så gör man det inte; antingen så går man och röstar eller så gör man det inte. Principen för en multinominal regression är emellertid densamma (se Garson för en närmare redogörelse).

I framställningen har vi betonat att logistik regression främst är en metod för hypotesprövning. Detta då vi menar att logistisk regression främst bör ses som en *faktororienterad* metod. Även om det naturligtvis går att använda metoden också för modellorienterade syften är detta emellertid mindre vanligt (och som ovan har konstaterats går det att ifrågasätta relevansen av de R^2 som finns till hands).

Vi har också varit måna om att man kan arbeta med logistisk regression utan att vara särskilt väl bevandrad i de matematiska principer som metoden bygger på. Visst är det bra om man vet vad den naturliga logaritmen är – men för att praktiskt arbeta med metoden behöver man inte förstå hur iterationsprocessen går till. (SPSS gör ju jobbet! Vad vi är intresserade av är att kunna *tolka* våra data!)

Men är då logistisk regression *alltid* att föredra framför OLS? Svaret blir: När vi inte tvingas göra våld på våra data är OLS en kraftfullare metod (risken för typ II-fel är mindre).¹⁵ Är man osäker på om kraven är uppfyllda bör man, menar vi, göra en logistisk regression.

För den som har fått mersmak på logistisk regression vill vi avsluta med en rekommendation för fortsatt läsning: Per Arne Tuftes *En intuitiv inneførelse i logistisk regresjon* är precis vad titeln lovar. På trevlig och lättbegriplig norska redogör författaren för metodens förtjänster utan att lägga för mycket vikt vid de tekniska detaljerna; det pedagogiska upplägget har gjort att Tuftes redogörelse faktiskt har fått fungera lite som vår förlaga!

¹⁵ Under FAQ skriver Garson, vilket förvånar oss en del, också att "[w]ith a binary dependent, it is impossible to meet the normality assumption, but if the split is not extreme (not 90:10 or worse), OLS regression will not return dramatically different substantive results".

6.1 Referenser

- Agresti, A. (1996): *An Introduction to Categorical Data Analysis*. New York: Wiley
- Aldridge, J.H. & Nelson, F.D. (1984): *Linear Probability, Logit and Probit Models*. Beverly Hills: Sage
- Aneshensel, C.S. (2002): *Theory-Based Data Analysis for the Social Sciences*. Thousand Oaks: Pine Forge Press
- Cabrera, A.F. (2002): Logistic Regression Analysis in Higher Education. An Applied Perspective. I Smart, J.C. (ed): *Higher Education. Handbook for the Study of Higher Education*. Vol 10. New York: Agathon Press
- Djurfeldt, G. & Barmark, M. (2009): *Statistisk verktygslåda – multivariat analys*. Studentlitteratur AB: Lund
- Garson, G.D. "Logistic Regression" from *Statnotes: Topics in Multivariate Analysis*. Retrieved 18/06/2009 from <http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm>
- Hagquist, C. & Stenbeck, M. (1996): *Goodness of Fit in Regression Analysis – R^2 and G^2 Reconsidered. Quality and Quantity*. Vol 32, No 3
- Hosmer & Lemshow (1989): *Applied Logistic Regression*. New York: Wiley
- Long, J.S. (1997): *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage
- Peng, C.Y.J., Lee, K.L. & Ingersoll, G.M. (2002): An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*. Vol 96, No 1
- Sirkin, R.M. (1997): *Statistics for the Social Sciences*. London: Sage
- Tufte, P.A. (2000): *En intuitiv innføring i logistisk regresjon*. Arbeidsnotatnr. 8-2000. Oslo: Statens institutt for forbruksforskning

PUBLIKATIONER UTGIVNA AV INSTITUTIONEN FÖR JOURNALISTIK, MEDIER OCH KOMMUNIKATION (JMG)

Institutionen för journalistik och masskommunikation (JMG) har två publikationsserier:

- JMG:s bokserie: Göteborgsstudier i journalistik och masskommunikation

- JMG:s arbetsrapportserie: Arbetsrapporter från Institutionen för journalistik och masskommunikation

Redaktörer för JMG:s bokserie och arbetsrapportserie är professorerna Kent Asp och Lennart Weibull. Publikationer i de två serierna säljs genom institutionens försorg. Beställningar kan ske på den talong som finns efter publikationsförteckningen.

Mellan 1991 och 1994 utgavs även en särskild rapportserie. Titlarna i den finns förtecknade separat.

JMG:s bokserie Göteborgsstudier i masskommunikation (1980 — 1990)

1.Lennart Weibull (1983) *Tidningsläsning i Sverige*. Stockholm, Liber. 521 sid.

2.Anders Ohlsson (1989) *Politiska nyheter till nytta och nöje*. Almqvist & Wiksell International, Stockholm. 350 sid.

3.Lennart Weibull, Karin Björkqvist (1989) *Dagspressen och dess läsare*. Stockholm, Almqvist&Wiksell International. (slut)

Göteborgsstudier i journalistik och masskommunikation (1990 —(Pris: 200 kr plus moms och porto)

1.Robert Burnett (1990) *Concentration and Diversity in the International Phonogram Industry*. (slut)

2.Roger Wallis (1991) *Internationalisation, Localisation & Integration*. 281 sid.

3.Ronny Severinsson (1994) *Tidningar i konkurrens. Dagstidningarnas agerande på lokala läsarmarknader i Västergötland 1950 - 1985*.

4.Bo Reimer (1994) *The Most Common of Practices. On Mass Media Use in Late Modernity*. Almqvist & Wiksell International. Stockholm. (slut)

5.Ulf Wallin (1995) *Vad fick vi veta? En studie i svenska nyhetsmediers rapportering åren före folkomröstningen om EU*.

6.Lennart Weibull och Charlotta Kratz (red) (1995) *Tidningsmiljöer. Dagstidningsläsning på 1990-talet*.

7.William Borden (1995) *Power Plays. A Comparison Between Swedish and American Press Policies*.

8.Admassu Tassew (1995) *Reporting a Pandemic. A Comparative Study of AIDS News Coverage in African and European Dailies*.

9.Monika Djerf Pierre (1996) *Gröna nyheter. Miljöjournalistiken i televisionens nyhetssändningar 1961 — 1994*.

10.Tomas Andersson Odén (1996) *Principer på pränt. En studie av redaktionella mål inom den svenska dagspressen*. Institutionen för journalistik och masskommunikation, Göteborgs universitet

11.Margareta Melin-Higgins (1996) *Pedagoger och spårhundar. En studie av svenska journalisters yrkesideal*. Institutionen för journalistik och masskommunikation, Göteborgs universitet.

12.Maria Elliot (1997) *Förtorendet för medierna. TV, radio och dagspress i allmänhetens ögon*. Institutionen för journalistik och masskommunikation. Göteborgs universitet.

13.Kent Asp, Bengt Johansson och Larsåke Larsson (1997) *Nära nyheter. Studier om kommunaljournalistik*. Institutionen för journalistik och masskommunikation. Göteborgs universitet.

14.Carlsson, Ulla (1998) *Frågan om en ny internationell informationsordning. En studie i internationell mediepolitik*. Institutionen för journalistik och masskommunikation, Göteborgs universitet.

15.Johansson, Bengt (1998) *Nyheter mitt ibland oss. Kommunala nyheter, personlig erfarenhet och lokal opinionsbildning*. Institutionen för journalistik och masskommunikation. Göteborgs universitet.

16.Wallin, Ulf (1998) *Sporten i spalterna. Sportjournalistikens utveckling i svensk dagspress under 100 år*. Institutionen för journalistik och masskommunikation. Göteborgs universitet.

17.Larsson, Larsåke (1998) *Nyheter i samspel. Studier i kommunaljournalistik. Institutionen för journalistik och masskommunikation*. Göteborgs universitet.

18.Weibull, Lennart och Ingela Wadbring /red/ (1998) *Publik och medier 1996/1997. Särtryck av artiklar om medier ur SOM-rapporterna nr 18 och 19*. Institutionen för journalistik och masskommunikation. Göteborgs universitet.

19.Lindstedt, Inger (1998) "Till de unga, till dem som ämna bliva tidningsmän". *Handböcker i journalistik*. Göteborgs universitet.

20.Dahlén, Peter (1999) *Från Vasaloppet till Sportextra. Radiosportens etablering och förgrening 1925 — 1995*. Stockholm: Stiftelsen Etermidierna i Sverige.

21.Löfgren Nilsson, Monica (1999) *På Bladet, Kuriren och Allehanda. Om ideal och organiseringsprinciper i den redaktionella vardagen*. Göteborgs universitet.

22.Wadbring, Ingela och Weibull, Lennart /red/ (2000) *Tryckt. 20 kapitel om dagstidningar i början av 2000-talet*.

23.Andersson Odén, Tomas (2000) *Skaraborgar'n och Spionen. Tidningar i Västra Götaland genom 250 år*.

23.Ghersetti, Marina (2000) *Sensationella berättelser. En studie av nyheter från Angola 1987 och om Prinsessan Diana 1997 i dagstidningar, radio och TV*.

JMG:s rapportserie (avslutad 1994)

Rapporter från Avdelningen för masskommunikation

- 1.Lennart Weibull (1984) *Dagspress och etermedier i Sverige 1979-1983*. (slut)
- 2.Jan Strid och Lennart Weibull (1984) *Läsvanor och Läsintrösen. 1979-1983*. (slut)
- 3.Ingela Strid och Lennart Weibull (1984) *Annonstidningsläsning i Sverige*.
- 4.Ronny Severinsson (1985) *Publiken möter kabel-TV*. (slut)
- 5.Lennart Weibull (1985) *Närradiolyssnandet i Stockholm 1984*. (slut)
- 6.Bo Reimer och Lennart Weibull (1985) *Dagspress på arbetsplatsen*.
- 7.Informatörer i Sverige: Arbetsmarknad Framtidsutsikter (1986) Utdrag ur utredningen om informationsutbildning i Sverige (*UHÄ-rapport 95:13*)
- 8.Ingela Strid och Lennart Weibull (1986) *Mediesverige 1986*.
- 9.Bo Reimer (1986) *Läsaren och tidningen*.
- 10.Rutger Lindahl (1986) *Nyheter om mordet på Olof Palme*.
- 11.Monika Djerf (1986) *Funktioner hos kabel-TV*.
- 12.Anders Ohlsson (1986) *Att svara eller inte svara - det är frågan*.
- 13.Ronny Severinsson (1987) *Den nya medieframtid - TV via satellit och kabel*.
- 14.Ingela Strid och Lennart Weibull (1988) *Mediesverige 1988*.
- 15.Ronny Severinsson (1989) *Agerande och utveckling inom dagspressen i Västergötland*.

Rapporter från Institutionen för journalistik och masskommunikation (avslutad 1994)

- 1.Emanuelsson, Eva & Karlsson, Kristina (1991) *Informatörer inför 90-talet*. Pris: 90 kronor.
- 2.Wallin, Ulf (1991) *Bilden av EG*. Pris: 120 kronor.
- 3.Wallin, Ulf (1992) *EG - Hot eller löfte? Argument i dagspress, radio och TV för och emot svenskt EG-medlemskap*. Pris: 100 kronor.
- 4.Flodin, Bertil (1993) *Samhällskommunikation under 80-talet. En kunskapsöversikt*. Pris: 150 kronor.
- 5.Severinsson, Ronny (red), (1995) - *Studier i medielandskapet. Ett urval av analyser genomförda på uppdrag av Pressutredningen -94*. Pris: 200 kronor.

Arbetsrapporter

Arbetsrapporter från Avdelningen för masskommunikation (1980 — 1990) (Pris: 60 kr + moms och porto)

- 1.Lennart Weibull (1984) *Trender i massmedieanvändningen*.
- 2.Bo Reimer (1985) *Tidningsläsning i södra Halland*.
- 3.Lennart Weibull (1985) *Läsvanor för svensk morgonpress 1979-1984*.
- 4.Ingela Strid, Lennart Weibull (1985) *Sport i medierna*.
- 5.Lennart Weibull (1985) *Masskommunikationen i Sverige. Några*

reflektioner om läget 1985.

- 6.Karl Erik Rosengren, Bo Reimer (1985) *Internaliserad kultur. Ett forskningsprogram om värden, individer och massmedier*.
- 7.Robert Burnett, Peter Esaiasson, Bo Reimer (1985) *Milestones in Mass Communication Research: Media Effects. A Review Essay*.
- 8.Bo Reimer (1985) *Values and the Choice of Measurement Technique. The Rating and Ranking of Postmaterialism*.
- 9.Lennart Weibull (1986) *Press, radio, TV och nya medier i Japan*.
- 10.Nyhetspridningen om mordet på statsminister Olof Palme. En sammanställning av Statistiska Centralbyråns intervjuundersökning 2-4 mars 1986.
- 11.Rune Hedberg och Lennart Weibull (1986) *Kassettidningsläsning på Gotland*.
- 12.Lennart Weibull (1986) *Massmediernas framtida utveckling*.
- 13.Ronny Severinsson (1986) *Tidningsläsning i Södermanland*.
- 14.Maria Elliot, Bo Reimer och Lennart Weibull (1986) *Tidningsläsning i Kalmar län*.
- 15.Lennart Weibull, Dagspresskollegiet (1986). *Läsvanor 1986*.
- 16.Bo Reimer och Karl Erik Rosengren (1986). *Maps of Culture: Macro and Micro*.
- 17.Maria Elliot (1986). *Allmänhetens syn på massmediernas trovärdighet*.
- 18.Lennart Weibull (1986) *Tendenser i svensk dagstidningsläsning*.
- 19.Bo Reimer (1986) *Social Space and the Structuring of Communication Processes*.
- 20.Lennart Weibull och Ronny Severinsson (1987) *Actions and Reactions - the Nordic Countries in Age of Satellite Broadcasting*.
- 21.Lennart Weibull (1987) *Massmediernas räckvidd 1985/86*.
- 22.Ronny Severinsson (1987) *Dagspressutvecklingen i Västergötland*.
- 23.Keith Roe (1987) *The Swedish Moral Panic over Video 1980-84*.
- 24.Bo Reimer (1987) *Dagens Nyheter i västra Sverige*.
- 25.Ann-Marie Hellerström (1987) *Lokal-TV- ett reellt alternativ till satellit-TV*.
- 26.Ronny Severinsson (1987) *Skaraborgspressens innehåll 1950-1985*.
- 27.Ronny Severinsson (1987) *Tidningsläsning i Essunga, Götene och Vara*.
- 28.Karin Björkqvist (1988) *Tidningsläsning i Gästrikland - presentation av en läsarundersökning*.
- 29.Bo Reimer (1988) *Reading Postmaterialism*.
- 30.Lennart Weibull (1988) *Publicistisk sed. Ett forskningsprogram om ett regelsystem och dess betydelse för svensk journalistik*.

31. Magnus Anshelm och Lennart Weibull (1988) *Läsvanestudien 1986. Huvudresultat och metodjämförelse.*
 32. Magnus Anshelm (1988) *Sverige - nu - SOM 86 - en jämförelse av två undersökningar.*
 33. Rutger Lindahl och Lennart Weibull (1988) *Palme i amerikanska ögon. En studie av den amerikanska pressens behandling av nyheten om mordet på statsminister Olof Palme 1986.*
 34. Magnus Anshelm (1988) *Massmediernas räckvidd 1986/87 - en analys av Mediebarometern.*
 35. Ingela Strid (1988) *Svenska folkets intressen 1973-1982. En sekundärbearbetning av Testologens mätningar av intressen.*
 36. Anders Ohlsson och Keith Roe (1988) *Mått på mätning - En förundersökning av fördelningsegenskaper och stabilitet hos linje- och kategoribaserade mått.*
 37. Britt Börjesson (1988) *Brott och publicitet. Kriminaljournalistik och pressetik under 1900-talet.*
 38. Lennart Weibull (1988) *Rapport från kontaktresa till Uruguay 10-22 april 1988.*
 39. Gunilla Jarlbro (1988) *En kvalitativ studie av hur människor upplever tidningsinnehåll.*
 40. Monica Löfgren (1988) *Kabel-TV i Göteborg.*
 41. Lennart Weibull (1988) *Lokal-TV via kabel. Synen på en ny kanal bland andra medier.*
 42. Karin Björkqvist (1988) *Tidningsläsning i Värmland. Presentation av en läsundersökning från våren 1977.*
 43. Keith Roe (1988) *Adolescents' VCR Use: How and Why.*
 44. Monika Djerf (1989) *Finns det funktionsdimensioner i medieanvändningen? En faktoranalys.*
 45. Karin Björkqvist (1989) *Det politiska tidningsvalet.*
 46. Magnus Anshelm och Jan Strid (1989) *Läsvanor och läsintressen.*
 47. Monika Djerf (1989) *Massmedier och beslutsfattare: en litteraturstudie.*
 48. Britt Börjesson (1989) *Pressens själusanering. Ett regelsystems framväxt.*
 49. Magnus Anshelm (1989) *Mediebarometern resultat från 1988/99.*
 50. Lennart Weibull (1989) *Rapport från en andra kontaktresa till Uruguay. Om massmedier, utbildning och forskning i social kommunikation samt Instituto Comunicacion y Desarrollo.*
 52. Monica Löfgren (1989) *Nyhetstidningen 8 sidor - innehåll, spridning och läsning.*
 53. Keith Roe (1989) *Notes on the Concept of Aggression and its (Mis) Use in Media Research.*
 54. Charlotta Kratz (1989) *Tidningsläsning i Kalmar och på Öland.*
 55. Catharina Kisch och Mikael Stoltz (1989) *City 103 och dess lyssnare - en undersökning om lyssnandet på svenska arbetsgivarreföreningens närradiostation i Göteborg.*
 56. Karin Björkqvist (1990) *Mediebarometern under åttiotalet.*
 57. Roger Wallis (1990) *Music, music everywhere, and so much of it the same.*
 58. Charlotta Lekvall och Patrik Vult von Steyern (1990) *City 103 och tre livsstilar.*
 59. Margareta Melin (1990) *Kommunaktuellt - studier av tidningens spridning och abonnenternas läsvanor.*
 60. Maria Elliot (1990) *Förtroendet för medierna.*
 61. Keith Roe (1990) *Never Has so Much Been Written by so Many about so Few, or, Why Youth Research?*
- Arbetsrapporter från Institutionen för journalistik och masskommunikation (1990 - (Pris: 80 kr + moms och porto)**
1. Ronny Severinsson (1990) *Massmedier, allmänhet och flyktingfrågor i Sverige och Sjöbo. En fallstudie på Sjöbo.*
 2. Ronny Severinsson (1990) *Tidningsläsning i Nordvästra Skåne.*
 3. Charlotta Kratz (1990) *Tidningsläsning i sydöstra Skåne. En presentation av en läsundersökning.*
 4. Bo Reimer (1990) *Kulturdiskussioner - Inlägg i tre eviga debatter.*
 5. Charlotta Kratz (1990) *Tidningsläsning i Jönköping. En presentation av en läsundersökning.*
 6. Charlotta Kratz (1991) *Verklighetsval och kapital. En studie av det ekonomiska och det kulturella kapitalets betydelse för läsningen av stockholms-tidningar utanför Stockholm.*
 7. Deanna Huthman (1991) *Dynamisk public relations. En explorativ studie av Public Relations i fyra företag.*
 8. Admassu Tassew (1991) *Public service broadcasting concept and practice: a preliminary overview of the ideals and some swedish experience.*
 9. Ronny Severinsson (1991) *Tidningsvanor i Gästrikland 1989. Presentation av en läsundersökning.*
 10. Lennart Weibull (1991) *Publik och medier i samspel. En analys gjord för MedieSverige 1991.*
 11. Monica Löfgren (1991) *Massmediernas räckvidd 1979 - 1990.*
 12. Lennart Weibull (1991) *Masskommunikation som ämnesområde. Ett försök till empirisk bestämning.*
 13. Charlotta Kratz (1991) *Tidningsläsning i Södra Halland 1990, jämförelser med situationen 1984.*
 14. Margareta Melin (1991) *Från kulturteori till journalistkultur. En litteraturoversikt över diskussionerna kring kulturbegreppet.*

- 15.Charlotta Kratz (1991) *Från icke-läsare till allätare. Fyra läsartyper i Stockholm och i landsorten.*
- 16.Kent Asp (1992) *Partiskheten i Sveriges Radios och TV4:s nyhetsprogram under 1991 års valrörelse.*
- 17.Ronny Severinsson (1992) *Med eller utan prenumeration i Eskilstuna och Katrineholm. Presentation av en läsa rundersökning från hösten 1990.*
- 18.Karin Björkqvist Hellingwerf (1992) *Mediebarometern 1979-1991.*
- 19.Karin Björkqvist Hellingwerf (1992) *Dagspress och medievanor. En analys av läsvanestudien 1991.*
- 20.Lennart Weibull och Karl Erik Rosengren (1992) *Swedes' view of the world. A Descriptive Report from the First Two BALTICOM Surveys.*
- 21.Monica Löfgren Nilsson (1992) *Kvinnligt, manligt, journalistiskt - journalisters syn på nyhetsvärdering.*
- 22.Monika Djerf-Pierre (1992) *A Toaster With Pictures. The Evolution of American Broadcasting 1921-1991.*
- 23.Charlotta Kratz (1992) *En fråga om smak. Om stabila och rörliga grupperns kulturella preferenser.*
- 24.Thomas Östberg (1992) *Sportjournalistik - en analys av fyra dagstidningar 1961 - 1991.*
- 25.Karin Björkqvist Hellingwerf (1993) *Mediebarometern 1979 - 1992.*
- 26.Carlin Nilsson, Åsa Widgren (1993) *Videotex - massmedium eller ej? En litteraturstudie och begreppsanalys.*
- 27.Margareta Melin (1993) *Var finns kvinnorna? En analys av manligt och kvinnligt i kurslitteraturen vid Journalisthögskolan i Göteborg.*
- 28.Karin Björkqvist Hellingwerf (1993) *Läsvanestudiens mått på läsvanor. En analys av två mått mått på läsvanor.*
- 29.Gunilla Jarlbro (1993) *HIV-aktuellt - En läsundersökning.*
- 30.Monica Löfgren Nilsson (1993) *Klimat och kön - Journalisters bedömningar av arbetsklimatet på nyhetsredaktioner.*
- 31.Bengt Carlsson, Inger Lindstedt, Lennart Weibull (1993) *Studieresa till amerikanska högskoleutbildningar i journalistik, medier och kommunikation - En personlig reserapport.*
- 32.Håkan Hvitfelt, Lennart Weibull (1993) *Pendeln har svängt - En reserapport från Kina.*
- 33.Ulf Wallin (1993) *Den franska folkomröstningen - En analys av rapporteringen om den franska folkomröstningen om Maastrichtfördraget i svensk press, radio och TV samt i övriga nordiska länders press.*
- 34.Karin Björkqvist Hellingwerf (1994) *Tidningsläsning i Luleå och Kiruna*
- 35.Tomas Andersson (1994) *Jönköpings-Posten och Smålandsposten - två småländska tidningsstrategier.*
- 36.Karin Björkqvist Hellingwerf (1994) *Mediebarometern 1979 - 1993.*
- 37.Bengt Johansson (1994) *Att studera massmediernas innehåll. En genomgång av innehållsstudier inom svensk forskning om journalistik, medier & kommunikation.*
- 38.Maria Edström, Maria Jacobson (1994) *Massmediernas enfaldiga typer. Kvinnor och män i mediebruset den 17 mars 1994.*
- 39.Ingela Wadbring, Lennart Weibull (1994) *På konsumenternas sida. En analys av hur konsumentfrågorna speglas i göteborgspresen.*
- 40.Bo Reimer (1994) *Kulturell identitet och massmedieanvändning på Åland.*
- 41.Olof Hultén, Carin Nilsson (1994) *Det svenska TV-utbudet 1987 och 1993. Undersökning av sändningstid och programkategorier.*
- 42.Kent Asp (1994) *Medieval 94'. Anförande vid Folkrorelsernas Medieforum torsdagen den 20 oktober 1994 (framfört i urval). Forskningsprogrammet Journalistik och demokrati - studier kring medie-kvalitet.*
- 43.Margareta Melin (1995) *Can Women Become Cowboys? The Importance of Journalist Education for the Professional Ideal Among Swedish Journalists.*
- 44.Margareta Melin (1995) *Female Educators and Male Craftsmen? The Professional Ideals Among Swedish Journalists.*
- 45.Ingela Wadbring (1995) *Sista ordet är inte sagt än! - en analys av lärarnas syn på jämställdhet vid journalistutbildningen i Göteborg.*
- 46.Karin Björkqvist Hellingwerf (1995) *Tidningsläsning i Gästrikland.*
- 47.Gunilla Jarlbro (1995) *Barn, ungdom och reklam.*
- 48.Karin Björkqvist Hellingwerf (1995) *Mediebarometern 1979 - 1994.*
- 49.Lennart Weibull (1995) *Journalister om Journalisten. Redovisning av en läsundersökning från hösten 1994.*
50. Kent Asp (1995) *Kommersialiserade TV-nyheter - på gott och ont. En jämförande undersökning av Rapport TV2 och Nyheterna TV4.*
51. Monica Löfgren Nilsson (1995). "Pennskaften" *Female Journalists in Sweden.*
- 52.Gunilla Jarlbro (1995) *Våldsbrott i svensk press. En jämförelse mellan åren 1983 och 1993.*
- 53.Jakob Bjur och Jörgen Thörnqvist (1995) *Samma röster med nya ansikten. En studie av Rapports innehålls-, form- och presentationsmässiga förändring över en tioårsperiod med tyngdpunkt på begriplighet.*
- 54.Dino Viscovi (1995) *Med Loket mot framtiden. Om unga arbetarklassmän, invandrare, EU och massmedier.*
- 55.Ingela Wadbring (1995) *Tidningsläsaren - vem är det? En analys av olika grupperns tidnings- användning.*
- 56.Marina Ghersetti (1996) *Jakten på den försvinnande publiken. Om besök och besökare på svenska biografier.*
- 57.Kent Asp (1996) *Sverigebildens i TV-nyheterna. En undersökning av Stockholmsdominansen i Rapport (1986 - 1995) och Nyheterna (1991 - 1995).*
- 58.Lennart Weibull (1996) *Bio, film och samhälle 1995. En analys av svenskarnas biobesök, filmpreferenser och syn på film och samhälle*
- 59.Birgitte Christiansen (1996) *IT-brug - en litteraturoversikt ud fra*

et humanistisk-samfundsvetenskapligt perspektiv.

60.Ylva Brune (1996) *Vålberg i nyheterna - en kamp mellan tolkningsmönster.*

61.Karin Björkqvist Hellingwerf (1996) *Mediebarometern 1979 - 1995.*

62. Ingela Wadbring (1996) *Den journalistiska vägen - vadan och varthän? En studie om kvinnliga och manliga journaliststudenter i Göteborg.*

63.Ylva Brune (1996) *Svenskar, invandrare och flyktingar i rubrikerna.*

64. André Jansson (1996) *TV-tittarna och det sociala rummet - En studie av den sociala positionens betydelse för kanalpreferenser och självuppfattning.*

65. Bo Reimer (1996) *Inte utan relevans - Om ungdomars användning av massmedier.*

66.Birgitte Christiansen (1996) *IT-brug i hverdagslivet- en kvalitativ interview-undersøgelse om ti menneskers computer- og Internetbrug i privatsfæren.*

67.Ebba Sundin (1997) *Barn och massmedier - En forskningsöversikt.*

68.Monica Löfgren Nilsson (1997) *Hon och han i journalistutbildningen - Ett pedagogiskt projekt.*

69.Bertil Flodin (1997) *Medborgarna och EU-informationen - En studie av inställning, aktivitet och kunskap hos svenska folket inför folkomröstningen 1994*

70.Maria Edström (1997) *Kön och journalistik - att utvidga medieverkligheten. En litteraturöversikt*

71.Larsåke Larsson (1997) *Forskningsöversikt - Kommunjournalistik.*

72.Torsten Malmström och Lennart Weibull (1997) *Snöstormen den 17 november 1995.*

73.Anna Maria Andersson (1997) *Vem är den svenske journalisten? - En studie av journalistkåren i yrkesliv och vardagsliv 1994 - 1995.*

74.André Jansson (1997) *Svensk dagstidningsläsning i förändring. Resultat från Riks-SOM-studien 1996.*

75.Britt-Marie Leivik Knowles (1997) *Inter-organisatorisk kommunikation mellan myndigheter och medier. Forskningsöversikt över myndigheters och mediers interaktion i samband med allvarliga samhällsstörningar.*

76.André Jansson (1997) *Sätt att se på TV. En kvantitativ problematisering av den selektiva TV-publiken.*

77.Sigurd Høst och Ronny Severinsson (1997) *Avisstrukturen i Norge og Sverige - 1960 til 1995. Arbeidsrapport nr. 1 fra prosjektet Norsk-svensk dagspresseutvikling.*

78.Anders Lithner och Ingela Wadbring (1997) *Ungdomars dagstidningsläsning.*

79.André Jansson (1997) *Tidningsvärden i morgon- och kvällspress.*

80.Tomas Andersson Odén (1997) *Pressen i Skaraborg. Tidningar och redaktörer från 1793 och framåt.*

81.Magnus Andersson (1997) *Populärmusik: såväl Wu-Tang Clan som*

Pat Bone.

82.Birgitte Christiansen och Annika Bergström (1998) *Tidningars symbol- & bruksvärde. Kvalitativa intervjuer med några av Göteborgs-Postens & Hallands Nyheter:s läsare.*

83.Erik Fjellman och Jan Sjögren (1998) *Nyhetsläsning på Internet - Svenska CNN från ett användarperspektiv.*

84.Ronny Severinsson /red/ (1998) *Lokalt innehåll i norska och svenska tidningar. Fyra redaktörer läser varandras tidningar.*

85.Anna Maria Andersson, Birgitte Christiansen & Karin Fogelberg (1998) *Vad tycker tittarna? En publikstudie av två underhållningsprogram.*

86.André Jansson (1998) *Dagstidningsläsning 1997. En tabellrapport från Dagspresskollegiet.*

87.Kent Asp (1988) *Medierna och giftutsläppet i Hallandsåsen.*

88.Ronny Severinsson (1988) *Marjasin i medierna. En innehållsanalys på uppdrag av Medieakademin.*

89.Annika Bergström (1988) *Tidningar, radio och TV i västra Småland.*

90.Annika Bergström (1999) *Internet i svenska hushåll hösten 1997.*

91.Lennart Weibull och Ingela Wadbring (1999) *De nya svenskarna möter svenska massmedier.*

92.Liselotte Englund (1999) *Katastroffjournalistikens dilemman. En forskningsöversikt med etiska och psykologiska perspektiv.*

93.Josefine Sternvik (1999) *Dagspressens ungdomssatsning.*

94.Maria Domellöf (1999) *Mycket nöje! Nöjesjournalistik i förändring.*

95.Annika Bergström (1999) *Internet i medielandskapet.*

96.Anna Maria Jönsson (2000) *Radio och TV i allmänhetens tjänst? Allmänhetens syn på public service-verksamheten.*

97.Liselotte Englund (2000) *Kvalitetseffektivitet i Public Service-radion. En kunskaps- och forskningsöversikt med förslag till indikatorer på radiokvalitet.*

98.Kent Asp, Bengt Johansson och Åsa Nilsson (2000) *Medievalsundersökningen 1998. Teknisk rapport.*

99.Magnus Fredrikson (2000) *Var skall jag kryssa? - En studie av RSV:s och dagspressens personvals information inför valet 1998.*

100.Stina Bengtsson (2000) *Personvalet och journalisterna - 16 svenska journalister om det första svenska personvalet 1998.*

1. Annika Bergström & Ingela Wadbring (2001) *Medierna i vardagslivet. En metodologisk skiss över hur det är möjligt att använda dagböcker som insamlingsmetod i medieforskningen.*
2. Kent Asp & Magnus Fredriksson (2001) *Bilavgaserna och den allmänna opinionen. Åsikter och handlingsberedskap.*
3. Tomas Andersson Odén (2002) *2001 års publicistiska bokslut. En rapport om 36 tidningars bokstäver - i siffror.*
4. Gabriella Sandstig (2002) *Organisation och kön. En strategisk litteraturöversikt av ämnesområdena organisation och kön applicerat på forskningsprojektet "Kvinnorna i journalistkulturen".*
5. Therese Eriksson (2002) *Gratistidningsmarknaden i Sverige 2001.*
6. Rudolf Antoni (2002) *Publikens publik. Aktuell forskning kring användning och värdering av medier.*
7. Rudolf Antoni & Therese Eriksson (2002) *Läsvanestudien. En tabellrapport.*
8. Gabriella Sandstig (2003) *Orons platser. En granskning av människors upplevelser av rädsla och otrygghet.*
9. Tomas Andersson Odén (2003) *2002 års Publicistiska bokslut. Del 1 Om tidningars redaktioner och innehåll.*
10. Ingela Wadbring (2003) *2002 års Publicistiska bokslut. Del 2 Om läsares och medarbetares syn på tidningar.*
11. Tomas Andersson Odén & Ingela Wadbring (2003) *Teknisk rapport för publicistiska bokslut.*
12. Mariann Björkemarken (2003) *Jönköpings Lokal-TV Förening. Utvecklingen av en lokal TV-verksamhet.*
13. Kent Asp (2003) *Medieval 2002 - partiskheten och valutgången. En studie av valrörelsens medialisering.*
14. Petra Sintorn (2003) *Publicistiska bokslut - hyllvärmare eller redskap för förändring?*
15. Oscar Westlund (2003) *Betala för nyheter på internet?*
16. Bengt Johansson, Lisa Henricsson & Annelie Karlsson (2003) *Hur farligt är Göteborg? Risker i GP:s lokalnyheter.*
17. Gabriella Sandstig (2004) *Att undersöka otrygghet - en metodstudie.*
18. Tomas Andersson Odén (2004) *2003 års Publicistiska bokslut. Del 1 Om tidningars redaktioner och innehåll.*
19. Ulrika Andersson (2004) *2003 års Publicistiska bokslut. Del 2 Sportens olika sidor: männens och de manliga sporternas revir.*
20. Josefine Sternvik (2004) *Ögonrörelser och dagstidningsläsning - en forskningsöversikt.*
21. Valpuri Mäkinen (2004) *En tidning för alla? Om invandrares användning av och attityder till medier i allmänhet och dagstidningar i synnerhet.*
22. Eva Berglie (2004) *Jakten på lycka. En teoriutvecklande studie av tredjepersonseffekten.*
23. Åsa Nilsson (2004) *Mediers räckvidd. En jämförelse av mätmetoder och undersökningar.*
24. Oscar Westlund (2004) *www.lokaltidning.se . Prenumeranterns attityder till sin lokaltidning och nättidningsläsning.*
25. Karin Fogelberg (2005) *Media Literacy. En diskussion om medieundervisning.*
26. Mira Öhlin (2005) *Fågel, fisk eller... Om klassresenärers medievanor.*
27. Tomas Andersson Odén (2005) *2004 års publicistiska bokslut. Om tidningars redaktioner och innehåll.*
28. Anna Maria Jönsson (2005) *Mångfalden i journalistikåren - studier av social sammansättning, samhällssyn och nyhets syn i den svenska journalistikåren.*
29. Tomas Andersson Odén (2005) *President Chavez i press, radio och TV. Mediebevakningen av folkomröstningskampanjen i Venezuela, augusti 2004.*
30. Ulrika Andersson (2005) *Journalister och deras läsare. En studie av tidningsjournalisters syn på läsare och läsarundersökningar.*
31. Ulrika Andersson, Ingela Wadbring (2005) *I gratistidningsland.*
32. Carl Bergholtz, Johan Olsson (2005) *Unga dagstidningsläsare i ny mediesituation.*
33. Åsa Nilsson, Lennart Weibull, Annette Hill (2005) *Synen på icke-fiktion i TV. Resultatredovisning.*
34. Ulrika Andersson (2005) *Nya svenskar och svenska medier.*
35. Ulf Wallin, Johannes Bjerling, Christoffer Lärkner (2006) *Barnen i nyheterna om tsunamin. En studie av mediernas rapportering under första månaden.*
36. Oscar Westlund (2006) *Media and Communication studies in Sweden. Disciplinary Boundary Construction - a theoretical contribution to Theory of Science.*

37. Oscar Westlund (2006) *Känslor av medieförtroende*.
38. Kent Asp (2006) *Journalistkårens partisympatier*.
39. Jonas Ohlsson (2006) *Partiernas press*
40. Ulrika Hedman (2006) *Har ni nåt webbigt? En fallstudie av Göteborgs-Postens flerkanalspublisering: Redaktionell organisering och arbetsrutiner, nyhetsvärde samt medielogik*
41. Anna Bolin (2006) *In whose interest? A journalists' view of their responsibilities and possibilities within the mainstream press in Sri Lanka*
42. Kent Asp (2006) *Rättvisa nyhetsmedier. Partiskheten under 2006 års medievalsvalrörelse*
43. Marina Ghersetti (2007) *Bilden av funktionshinder. En studie av nyheter i Sveriges Television*
44. Malin Dellgran (2007) *Konsten att förhandla. Om journalister i en kommersiell tv-kultur*
45. Johannes Bjerling (2007) *Partiledarutfrågningarna i SVT 2006*
46. Ulrika Andersson (2007) *Fokus på unga vuxna. Sociala förändringar och växande medieutbud skapar nya medievanor bland unga*
47. Oscar Westlund (2007) *Mobiltelefonianvändning. En forskningsöversikt*
48. Jonas Ohlsson (2007) *Göteborgsakademiker 2006. Medieinnehav och medieanvändning bland unga högutbildade*
49. Ingela Wadbring (2007) *Från annonsblad till tidningar. Decenniers utveckling av lokala gratistidningar*.
50. Oscar Westlund (2007) *Mobiltelefonen som multimediet och nyhetsmedium. Ett användarperspektiv*.
51. Gabriella Sandstig (2007) *Förtroendet för polisen. En analys av medborgarnas syn på polisen och dess roll i samhället 1986 - 2006*.
52. Sandra Byström, Björn Danielsson, Bengt Johansson (2007) *Regionala TV-nyheter i medielogikens tid. En studie av Västnyttens bevakning av Västra Götalandsregionen*.
53. Mathias A Färdigh (2008) *Mediesystem i Europa. En studie av de europeiska ländernas mediesystem utifrån ett användarperspektiv*.
54. Tomas Andersson Odén, Johannes Bjerling (2008) *A-press med borgerliga ägare*.
55. Therese Blomqvist, Maja Norin (2009) *"Sjuning, åttning, nining Tidning!" -Förskolebarns föreställningar om dagstidningen*
56. Ulrika Hedman (2009) *Läsarmedverkan: Lönande logiskt lockbete. Nättidningarnas användarskapade innehåll ur ett redaktionellt ledningsperspektiv*.
57. Magnus Ekström, Björn Granstrand (2009) *Dagstidningen på 3,5 tum? En nutids- och framtidsstudie av svensk dagspress förhållande till mobila tjänster*.
58. Johan Cedersjö, Roger Gustavsson (2009) *Hot eller möjlighet? En kvantitativ studie av svenska landsortsbaserade dagspressjournalisters inställning till användarskapat innehåll*.
59. Marina Ghersetti (2009) *Personer med invandrarbakgrund i SVTs nyhetsprogram. En innehållsanalys av Rapport, Aktuellt, Sportnytt och Kulturnyheter 2001*.
60. Marina Ghersetti (2009) *Terrorattackerna i USA. Sveriges Radios rapportering om den 11 september 2001*.
61. Monica Löfgren Nilsson (2009) *Kvinnor, män och nyheter i televisionen. En innehållsstudie av ämnen och källor under 45 år*.
62. Johannes Bjerling, Jonas Ohlsson (2010) *En introduktion till logistisk regressionsanalys*.

Vissa av rapporterna finns tillgängliga på JMGs hemsida. **www.jmg.gu.se**

**Institutionen för
journalistik, medier och kommunikation**

Box 710
405 30 GÖTEBORG

TEL 031 - 786 4976
FAX 031 - 786 45 54
e-post majken.johansson@jmg.gu.se